

The red planet's big picture

Let's not get carried away with the images from NASA's Spirit rover, nor despair of Britain's silent Beagle 2. The best way to explore Mars is through a programme that builds patiently on each mission's successes and failures.

With two probes safely arrived at Mars, another en route, and just one apparent casualty, space scientists can count themselves lucky. The planet known for wrecking spacecraft seems to have been more hospitable this time.

The familiarity of the pictures beamed back by NASA's Spirit rover from the martian surface belies their rarity. Until now, only three landers had made it to the surface of Mars — two Viking craft in 1976 and the Mars Pathfinder with its mini-rover in 1997, which was little more than a technology test bed.

Spirit and its soon-to-arrive twin, Opportunity, will raise the game to the next level (see page 89). Their cameras and spectrometers are several generations better than those carried by the Viking landers. They can roam around to interesting rock outcrops, like a field geologist. And now that Europe's Mars Express has joined NASA's Mars Global Surveyor and Mars Odyssey in orbit, Mars exploration is poised to enter a new, more sophisticated phase.

After Spirit's landing, Charles Elachi, director of the Jet Propulsion Laboratory in Pasadena, California, which built the Mars rovers, gave special thanks to an influential California congressman and a former White House budget official who lent key support to the Mars programme several years ago. It was a fitting gesture: without a significant budget increase, the lab might have repeated the débâcle of 1999, in which two of its Mars probes were lost.

The new rovers cost a total of \$820 million, more than twice the price tag of the 1999 missions. Not only do these rovers carry more-capable scientific instruments, they also possess additional technologies to improve reliability — or at least provide better information on any failures that may occur. This time, Spirit was in radio contact during much of its descent to Mars, so ground controllers could monitor key events such as the release of its parachutes.

Beagle 2 — which as *Nature* went to press was the only missing

member of the current Mars armada — had no communications capability between leaving its mother ship, Mars Express, and arriving on the planet's surface. It deserves high marks for even getting to the launch pad on a shoestring budget estimated at just \$60 million. But Beagle's British managers may wish they hadn't skimped quite so much. If Mars Express never picks up a signal, we may never know for certain what went wrong — whether the lander's parachutes or airbags failed catastrophically, or whether an unfortunately located crater rim blocked transmissions from an otherwise healthy spacecraft. Without such knowledge, it is hard to justify investing in a speculative Beagle 3 mission.

NASA, on the other hand, is beginning to reap the benefits of an integrated Mars programme, where each mission builds on the last. The agency's two Mars orbiters played key roles in scouting for Spirit's landing place, as well as communicating with the probe during landing. NASA's science plan is also patient and methodical, laying the geological groundwork by looking for evidence of past water, rather than rushing straight to the search for life.

The European Space Agency (ESA) also seems to be thinking long-term. Mars Express should produce two extremely valuable data sets — medium-resolution imagery of the whole planet, and an inventory of water and ice below the surface — that will inform more detailed future studies. Despite the greater media attention garnered by Beagle 2, Mars Express is the real milestone for European planetary science.

What is the right amount to spend on a planetary mission? That remains a tough question. But with successes such as last week's, ESA and NASA are learning to answer it with more assurance, and to estimate better what it takes to operate reliably on another planet.

Congratulations, then, to scientists and mission controllers on both sides of the Atlantic. It's good to be back on Mars. ■

China and SARS: could do better

Has the world's most populous nation learnt the lessons of last year's SARS outbreak? Only partially, it seems.

Some 10,000 palm civets in the markets in and around the Chinese city of Guangzhou in Guangdong province are due to be slaughtered. The animals, which are sold for food, will be dispatched because they carry the virus that causes severe acute respiratory syndrome (SARS), which has turned up in a resident of the region — the first case, apart from those caused by laboratory contamination, since last year's global outbreak (see page 89).

The slaughter plan and the quick reporting to the World Health Organization (WHO) of the infected patient are a welcome sign that China is now taking its public-health responsibilities seriously. This is in marked contrast to the tardy and secretive response to the previous outbreak, which first emerged in rural areas of Guangdong province in late 2002. However, the case also exposes shortcomings in China's system for monitoring and responding to infectious disease.

There is no evidence that slaughtering civets will prevent further

human cases of SARS. The natural animal reservoir for the SARS virus remains unknown — civets may be infected in markets from some other animal, and we don't even know whether they can pass the virus to people.

Still, given that the animals are destined for slaughter anyway, one could argue that disposing of them quickly is a reasonable precaution. The real question, however, is why they are on sale in the first place. At the height of the SARS outbreak, China closed its wildlife markets, only to lift the ban in August. That relaxation was unwise.

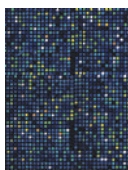
Questions also remain about the adequacy of China's system of monitoring for SARS and other emerging diseases. Away from major cities, many clinics — especially those specializing in traditional medicine, which are the first stop for most of the population — are not integrated into the disease-surveillance system. If China is to protect its people from re-runs of last year's SARS nightmare, this needs to change. ■





Herbal abuse

Diet pill faces US ban over potentially fatal side effects
p90



Array of hope

Geneticists seek standards for experimental data
p91



Leading role

Canada's prime minister sets science centre-stage
p91



Not guilty

Espionage charges against Russian physicist are dropped
p92

Rover barks before Beagle as Mars success lifts NASA's spirits

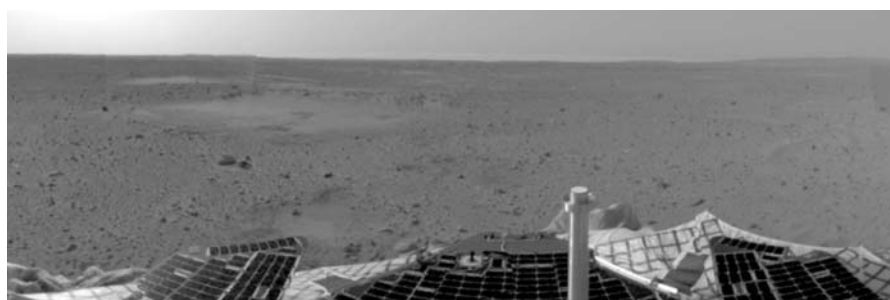
Tony Reichhardt, Washington

After landing on Mars on 3 January, NASA's Spirit rover is now flexing its mechanical limbs and surveying its surroundings before heading out to explore. Scientists and engineers at NASA's Jet Propulsion Laboratory (JPL) in Pasadena, California, who operate the desk-sized robot, are itching to investigate the various features seen in the pictures of the martian surface returned by the rover.

Based on those preliminary images, Steve Squyres, a planetary scientist at Cornell University in Ithaca, New York, and principal investigator for Spirit's scientific instruments, speculates that the rover has landed in a field of small craters at the centre of Gusev Crater, which is thought to be an ancient lake bed.

The terrain around the landing site is pocked with shallow depressions, which Squyres says may be secondary impact craters caused by material falling back to the surface after being thrown up by a larger impact. One such pit, which the Spirit team calls 'Sleepy Hollow', is likely to be the robot's first destination, as it looks to contain a rock outcrop that could offer insight into sedimentary history.

The rover had the good fortune to land in what Squyres called "the place where we absolutely wanted to be". The site is relatively free of boulders that would impede the robot. So mission controllers can bask in the



Land ahoy! Spirit's early view of Mars shows the small pit nicknamed 'Sleepy Hollow' centre left.

smooth, early operation of what was seen as a make-or-break mission for the JPL, which suffered embarrassing back-to-back failures in 1999 (see *Nature* 423, 477; 2003).

Europe's first planetary orbiter, Mars Express, is also off to a promising start, with two successful orbit manoeuvres already under its belt, and another scheduled for this week to bring it closer to the planet's surface. Agustin Chicarro, the European Space Agency's project scientist for the mission, says that data from the onboard instruments should begin flowing in late January.

One of the spacecraft's key instruments, a ground-penetrating radar that aims to map water and ice to depths of several kilometres below the martian surface, will be deployed in late February. By April, says Chicarro, Mars

Express should settle into a schedule of routine scientific observations.

James Head, a planetary geologist at Brown University in Providence, Rhode Island, says that if the Mars Express radar works as advertised, it could shed light on what happened to the large reservoirs of water thought to have existed in the planet's distant past.

A second NASA rover, called Opportunity, is set to reach Mars on 24 January. That leaves the British-built Beagle 2 lander as the only troubled mission in the current invasion of Mars (see *Nature* 427, 5; 2004). Project managers hope that Mars Express can make contact with Beagle 2 sometime after 7 January. High-resolution cameras on NASA's orbiting Mars Global Surveyor have also been enlisted to take close-ups of the lost craft's landing site. ■

Swift response greets return of SARS in China

David Cyranoski, Tokyo

China is rushing to quell public fears after confirmation on 5 January of the first naturally acquired case of severe acute respiratory syndrome (SARS) since the disease was contained in July last year.

The patient, a television producer from the southern province of Guangdong, became ill on 16 December. He was isolated in hospital in the provincial capital of Guangzhou, and is now making a good recovery. The quarantine has been extended to 81 of his contacts, all of whom have remained healthy. "The system of identification and contact-tracing seems to be

working," says Julie Hall, the Communicable Disease Surveillance and Response Coordinator for China at the World Health Organization (WHO).

The case was confirmed by tests carried out at two laboratories in Hong Kong and one in Beijing, which found a sharp rise in the patient's production of antibodies to the SARS virus. Researchers at the University of Hong Kong also say that the patient's virus is a close genetic match to samples taken from masked palm civets (*Paguma larvata*) and other animals in the markets in and around Guangzhou. On the basis of these results,

provincial officials plan to kill some 10,000 civets currently held for sale for food.

But experts are questioning this strategy. "Raccoon dogs, badgers, ferrets, house cats and rodents can all carry the SARS virus," says Albert Osterhaus, a virologist at Erasmus University in Rotterdam, the Netherlands. "A more comprehensive approach is needed."

Meanwhile, concern about the resurgence of SARS has spread to the Philippines, where a domestic maid working in Hong Kong who returned home for a holiday has shown suspicious symptoms. Tests for SARS were being conducted as *Nature* went to press. ■

Yucca review-board head quits in face of impartiality charges

Geoff Brumfiel, Washington

The embattled head of a review board for the proposed Yucca Mountain nuclear-waste repository has resigned. His critics had charged that he was too closely tied to the nuclear industry.

Michael Corradini, a nuclear engineer at the University of Wisconsin at Madison, announced his resignation from the Department of Energy's Nuclear Waste Technical Review Board on 30 December. His public support for the Yucca Mountain Project and his pursuit of nuclear research had led project opponents, including all five members of Nevada's congressional delegation, to question his impartiality and call for his resignation.

The review board is set up as an independent body, and has been charged with evaluating the energy department's licence application for the Yucca project (see *Nature* 418, 262; 2002). It has long been considered an impartial judge of technical questions, says Judy Treichel, executive director of the Las Vegas-based Nevada Nuclear Waste Taskforce, which opposes the project.

Treichel claims that this record of impartiality was threatened in June 2002, when the Bush administration appointed Corradini as chair. Corradini has testified to Congress in support of the project. He also continued to receive funding from the energy department for his research on nuclear power — something other board members say they have stopped doing. The congressional delegation sent a letter to President George Bush on 25 February 2003 calling for Corradini's resignation. Faced with such local opposition, the nine other board members wrote to Corradini in May, asking him to go.

Corradini at first refused to step down, saying that he would not give in to "political pressure". Nor did he back away from support of the project, co-authoring a newspaper article last October that stated that "Yucca Mountain is doable".

"It was a bit like a judge making a statement about the verdict before the trial has begun," says Norman Christensen, an ecologist at Duke University in North Carolina and a board member.

Corradini disputes that. "I don't think that me stating a fact makes me unacceptable, and I also don't think that makes me lack objectivity," he says.

Nonetheless, he concedes that opposition to his appointment had made it difficult for the committee to function, and so offered his resignation. ■

Safety concerns prompt US ban on dietary supplement

Jonathan Knight, San Francisco

The US government is to ban the popular dietary supplement ephedra because of mounting evidence of its toxicity. The ban has been welcomed by public-health advocates, but some critics have branded it an overreaction to adverse publicity about the drug.

Some 12 million Americans take diet pills or consume energy drinks laced with *Ephedra sinica*, a Chinese medicinal herb containing stimulants known as ephedrine alkaloids. But the 1994 Dietary Supplement Health and Education Act prevents the Food and Drug Administration (FDA) from enforcing safety studies for herbal supplements, as long as the product does not claim to cure any disease. Dozens of US companies sell ephedra-containing products.

But in a press release on 30 December, the FDA said that ephedra had been "conclusively linked to significant and substantial adverse health effects like heart problems and strokes". Once the agency's ruling is published, manufacturers will have 60 days to take their products off the market.

Since 1994, the FDA has received hundreds of complaints of adverse reactions to ephedra, ranging in severity from heartburn to death. But because the herb is so popular, it has been hard to determine which of the many effects reported were actually caused by the drug.

"Heart attacks happen anyway, and millions of people take ephedra, so these two universes will intersect," says Paul Shekelle, director of the Southern California Evidence-based Practice Center in Santa Monica. The centre is run jointly by five medical research centres and the Rand Corporation, a non-profit consultancy that assesses technical problems.

In 2001 the FDA asked the centre to analyse all available data on ephedra. Its conclusions, published last March (P. G. Shekelle *et al. J. Am. Med. Assoc.* 289, 1537–1545; 2003), linked the substance to minor side effects, but was less clear-cut on its connection with more serious effects such as heart attacks.

Of 18,000 adverse-event reports made either to the FDA or to an ephedra manufacturer during 2000, only 21 contained sufficient information to conclude a probable role for ephedra, and only two of these were reports of deaths. "This was not a slam dunk in either direction," Shekelle says.



Fitness factor: ephedra was linked to the death of Orioles pitcher Steve Bechler in February 2003.

Nevertheless, several high-profile deaths have raised public awareness of ephedra's potential risks. Last February, baseball pitcher Steve Bechler died of heat stroke while training for the Baltimore Orioles in Florida. The coroner concluded that ephedra, which Bechler had been taking for a weight problem, was a significant factor. And the deaths of Illinois high-school athlete Sean Riggins in 2002 and Minnesota Vikings tackle Korey Stringer in 2001 were also linked to ephedra. None of these was included in the data Shekelle reviewed.

Critics suspect that this adverse publicity has been a key factor leading to the ban. "We should rely more on scientific data than anecdotes in setting public-health policy," says Carol Boozar, an obesity researcher at St Luke's-Roosevelt Hospital Center in New York. And Steven Karch, a cardiac pathologist in San Francisco who studies stimulant abuse, says he has found no evidence that ephedra has played a role in any local death in the past decade. "We just don't see it," he says.

The FDA has not yet detailed its case supporting the ban, but says it will publish the ruling within a few weeks. Meanwhile, consumers are stocking up and there have been reports of health stores selling out of ephedra as fast as they can fill the shelves. ■

Geneticists chip away at unruly data

Erika Check, Washington

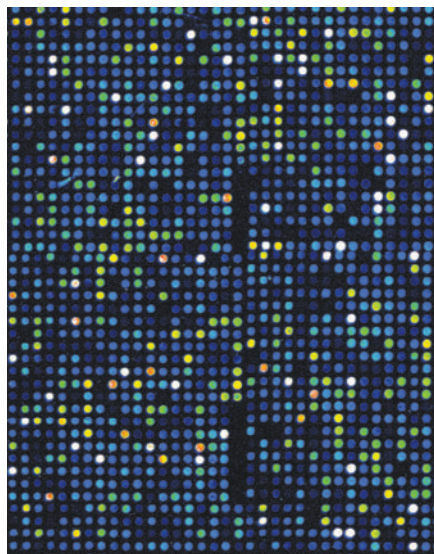
Researchers in genetics trying to share their results face an uphill struggle. The wide range of experimental methods that they use makes it fiendishly hard to compare data. But a consortium of companies and academics is now working with the US government to tackle the problem by drawing up a better standards system.

The External RNA Controls Consortium (ERCC) says that such standards will allow geneticists to reach more accurate conclusions, and to compare their data with the findings of others.

"There is definitely a need for standardization," says Gavin Sherlock, a geneticist at Stanford University in California. "It's one of the biggest problems we have — there is no way of connecting the dots between two platforms very well."

The bulk of genetics research involves characterizing the genes expressed by a particular cell or tissue under specific experimental conditions. But there are many ways to do this, and researchers say that they can get different results from the same experiment depending on the method they use. They can even have problems duplicating their own results from the same technology.

The ERCC is addressing the issue by setting up standard controls for checking the performance of experiments. The group is drawing up an extensive list of genetic sequences that won't normally turn up in most studies. Pieces of genetic material that



Sharing nature: fresh standards should make it easier to compare results from microarrays.

correspond to these sequences will then be manufactured for use as controls, allowing geneticists to test how well they are detected in their laboratories. Researchers will then be able to compare the results of different experiments with more confidence, consortium members say.

The ERCC held two workshops last year, and gene-chip manufacturers such as Affymetrix, based in Santa Clara, California, are already gearing up to produce the materials that the plan requires. Marc Salit of the

National Institute of Standards and Technology, which helped to organize the workshops, says that a non-profit means will probably be established to distribute the controls. He says that in the most optimistic scenario, the materials could be available within a year to 18 months.

The ERCC was set up in the wake of a standardization effort led by the Microarray Gene Expression Data Society, a group of scientists that has drawn up a common language for comparing gene-expression experiments and which is now devising data formats to help researchers to compare their results (see *Nature* **415**, 946; 2002).

Janet Warrington, vice-president of Affymetrix, says that one of the ERCC's goals is to move genetics closer to the clinic. Making sophisticated genetic analysis accessible to physicians will require much smoother standardization than now exists, she says: "If you're going to develop something for clinical application, you really want to look into developing standards."

Researchers say that the ERCC's work will be a big help, but will not entirely solve the problem of standardization. For instance, says John Quackenbush, a bioinformatician at The Institute for Genomic Research in Rockville, Maryland, the consortium's efforts are primarily focused on experiments in humans, so research in other organisms could be left out in the cold. "Will it solve all the problems? Absolutely not," Quackenbush says. "But it's a very important first step." ■

Canadian prime minister makes science a priority

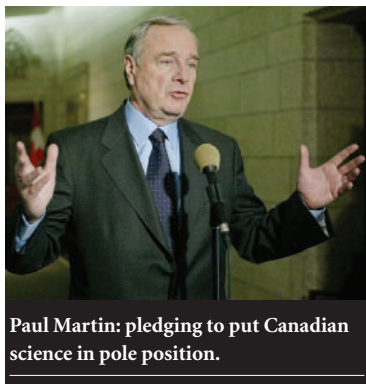
David Spurgeon, Montreal

Scientists in Canada are feeling a rush of optimism over the words and actions of the new prime minister, Paul Martin, who says that if he was born again, he'd like to come back as a basic researcher.

Martin, who succeeded fellow Liberal Jean Chretien

on 12 December, first made the remark in a recorded message to his supporters. The comment was then repeated in a November newspaper interview, in which Martin vowed to make science and technology a central pillar of government policy. "I'm making it a fundamental tenet of my government," he told the *Toronto Star*.

On the day he took office, Martin named



Paul Martin: pledging to put Canadian science in pole position.

Arthur Carty, president of the National Research Council Canada, the country's main research and development agency, as his full-time science adviser — the first such appointment in Canada for 30 years. Martin also says that he plans to chair a new cabinet committee to set scientific priorities, and will work to revitalize the Advisory Council on

Science and Technology, an external panel that has had a low profile in recent years. In addition, he has appointed seasoned politician Joe Fontana as parliamentary secretary for science and small business.

John Polanyi, a chemist at the University of Toronto who won the Nobel prize in 1986 for his work on reaction dynamics, says that Martin's first month in office is "a hopeful

moment in the history of Canadian science". But he cautions that Martin will have to resist pressure to over-commercialize research, especially at universities. "Our scientists are becoming less effective because of baroque schemes for milking science of its wealth-creating aspect," he complains.

Martin was finance minister when Chretien's government reversed cuts that had hampered the three federal agencies that fund most university research, and found billions of dollars for such initiatives as the Canada Foundation for Innovation.

In a speech in Montreal in October, Martin hinted at a more commercial approach to research policy, calling for "a fundamental change in the way that our research institutions assess the economic potential of their discoveries". But he later added that "basic research that does nothing else but expand the sum of human knowledge is sufficient reward in itself for the expenditure that government puts in." ■

Animal-rights groups take legal action to block primate centre

London Plans by the University of Cambridge to build a centre for brain research on primates face a fresh challenge, in the form of a lawsuit from two antivivisection groups.

In November, UK deputy prime minister John Prescott overruled South Cambridgeshire council's decision to refuse planning permission for the proposed Centre for Behavioural Neuroscience (see *Nature* **426**, 376; 2003). But Animal Aid and the National Anti-Vivisection Society are now challenging Prescott's ruling in the High Court. The groups point out that an independent planning inspector found that the university had failed to demonstrate that there was a "national need" for the lab.

The centre, first proposed in 1998, has become a flashpoint for Britain's bitter debate over animal experimentation. Over the lengthy planning process, the projected costs have risen by some £8 million (US\$14 million) to £32 million — giving the university authorities a headache even if the latest legal challenge proves unsuccessful. A university spokesman says that discussions with funding bodies, including the Medical Research Council and the Wellcome Trust, a research-funding charity, are continuing.

Japan agrees to hand over espionage suspect

Tokyo Japan is set to extradite a molecular biologist to the United States nearly three years after Washington charged him with stealing genetic research material, Japanese newspapers reported last week.

The Japanese justice minister, Daizo Nozawa, will begin procedures this month to transfer Takashi Okamoto, indicted on industrial-espionage charges in May 2001,

Russia drops spy charge

Moscow A jury in Siberia last week acquitted physicist Valentin Danilov (pictured) of charges of spying for China. Danilov — previously a professor at Krasnoyarsk State Technical University — had spent 19 months in jail after being accused by the Federal Security Service (FSB) of selling classified space-technology information to China (see *Nature* **424**, 477; 2003). Danilov's defence, publicly supported by prominent physicists including 2003 Nobel laureate Vitaly Ginzburg, was that the information came from open sources, including published journals.

Danilov's trial is one of a series brought against researchers by the FSB. Trial by jury is currently a rarity in Russia having been used in only 9 of the nation's 89 regions in the past ten years.



I. NAYMUSHIN/REUTERS

to the United States, according to the newspapers *Yomiuri Shimbun* and *Kyodo News*. The Tokyo High Court is expected to approve Okamoto's extradition after a two-month examination of the case.

Okamoto, a former laboratory team leader at the Institute of Physical and Chemical Research, or RIKEN, in Japan, was charged with violating the US 1996 Economic Espionage Act, which targets the theft of commercially valuable trade secrets to benefit a foreign government — the first time that the United States has invoked the relevant clause in the act. The charges came after the Cleveland Clinic Foundation in Ohio, where Okamoto studied Alzheimer's disease before returning to Japan in 1999, alleged that he had stolen DNA samples and cell lines. Economic espionage charges carry a stiff penalty of up to 15 years in prison, \$500,000 in fines, or both.

Stem-cell centres get green light in Spain

Munich Spain is to boost its commitment to research on human stem cells — an unusual move for a predominantly Roman Catholic country — by creating two centres to bolster research in regenerative medicine.

One facility, a National Center for Organ and Tissue Transplant and Regenerative Medicine, will be based in either Madrid or Barcelona, and will be networked with smaller centres throughout Spain.

The second, based in San Sebastian, will provide one of the world's largest zebrafish facilities for screening small molecules and potential drugs. Researchers will also use the fish to try to work out the genetic basis of tissue regeneration. "Zebrafish are unusual in that some tissues, like the heart, can regenerate if damaged," says Juan Carlos Izpisua Belmonte, a gene-expression expert at the Salk Institute for Biological Studies in La Jolla, California, who will head the centre.

The centres' backers say that they will supplement Spain's modest capability in fundamental stem-cell research.

Indian rocket to launch collaboration with Israel

New Delhi India will put a US\$14-million Israeli astronomical payload into space in 2005, in a sign of growing scientific collaboration between the two nations.

The Israeli payload, named TAUVEV (Tel Aviv University Ultra Violet Experiment), consists of three wide-field ultraviolet telescopes that will be used to study black holes and the formation of stars.

The deal was signed in Bangalore on 25 December by G. Madhavan Nair, chairman of the Indian Space Research Organisation (ISRO), and his Israeli counterpart, Aby Har-Even. Israel turned to India after an initial deal with Russia fell through. ISRO officials said that TAUVEV will ride piggy-back with ISRO's own communications payload on the geostationary satellite GSAT-4, scheduled for launch in September 2005.

Israeli science minister Eliezer Sandberg also said that Israel may participate in India's planned mission to the Moon in 2008.

Institute opens door on self-organizing systems

Munich The new year sees the launch of the Frankfurt Institute for Advanced Studies (FIAS) — an independent centre for basic multidisciplinary research into complex systems.

The FIAS, based at the Johann Wolfgang Goethe University in Frankfurt, will give biologists, neuroscientists, physicists, chemists and computer scientists the chance to conduct research on self-organizing systems. The first visiting fellows will be selected in the next few months from about 70 applicants, many of whom are from Russia and eastern Europe.

Wolf Singer, director of the Max Planck Institute for Brain Research in Frankfurt, and a founding director of the new institute, says that it will study the self-organizing principles behind complex systems, and the conceptual similarities between them.



A Japanese researcher allegedly stole genetic material from the Cleveland Clinic Foundation.

D. MAXWELL/AFP

Dreaming on the Danube

Can Budapest regain its status as one of Europe's scientific hubs? Perhaps, if the generation gap between Soviet-era scientists and young, westward-looking researchers can be bridged. Quirin Schiermeier reports.

Budapest's intellectual heart beats in the city's historic castle district, in a beautiful baroque building on Trinity Square, just opposite the famous Matthias church. The former city hall of Buda, which sits on a hillside overlooking its twin city, Pest, now plays host to the Collegium Budapest. This haven for theorists has a rarefied, cloistered atmosphere — as befits an institute established in an unashamed attempt to emulate the Princeton Institute for Advanced Study in New Jersey.

That's an ambitious aim: launched in 1930, the Princeton institute's early fellows included such luminaries as Albert Einstein and the logician Kurt Gödel. The Collegium Budapest can't yet claim to compete in the same league, but after just a decade in operation, it has established itself as one of the most prestigious academic addresses in Eastern Europe. With five permanent research fellows plus a roster of 20 or so visiting scientists, the Collegium is carving out a niche for theoretical work in disciplines from physics, through linguistics, to evolutionary biology.

Hungarian scientists hope that the Collegium's success can serve as a springboard to reviving their country's past scientific glories. Budapest, they argue, is ideally placed to lead attempts to re-establish a cosmopolitan culture of science among the countries of the former Eastern bloc. "Hungary has left important fingerprints on science in the past, and I am optimistic that Budapest will be able to regain its traditional role of a premium intellectual centre for central and eastern Europe," says neuroscientist Sylvester Vizi, who in May 2002 became president of the Hungarian Academy of Sciences.

The timing is right: for Hungary and nine other countries in the region, full membership of the European Union (EU) lies just months away. But many of those who want to reach out to the West point to a nagging obstacle: the old guard of Soviet-era scientists who still dominate the Hungarian scientific establishment.

No one can doubt Hungary's rich scientific pedigree. The country has produced

more Nobel laureates per capita than any other. Some say that Hungarians are predisposed towards logical thinking by the struggle to master their notoriously complex language. But whatever the reason, that language has filled the air at some of the most pivotal moments in twentieth-century science. Many of the discussions in the run-up to the explosion of the first atom bomb, for instance, were conducted in Hungarian — the presence on the Manhattan Project of Leo Szilard, Eugene Wigner, John von Neumann and Edward Teller saw to that.

"We always had strong links to the West, but the relation is lopsided," says Imre Kondor, a physicist at Budapest's Eötvös Loránd University and rector of the Collegium



Niche market: the Collegium Budapest is gaining an international reputation for theoretical research.

Budapest. While young Hungarian scientific talent continues to enrich leading labs around the world, few foreign researchers visit Hungary, even for short sabbaticals — the facilities just aren't attractive enough.

Thanks to funding from the EU and individual European governments, and by concentrating on inexpensive theoretical research, the Collegium has made itself an exception to this rule. Tamar Gendler, a philosopher and psychologist from Cornell University in Ithaca, New York, is one of 18 foreign fellows currently working at the institute. Gendler was attracted by "the ridiculous beauty of the guest house in walking distance to my office and the inspiring discussions with co-fellows", and is using her sabbatical to write a book on the neuronal basis of imagination, self-deception and 'fictional' emotions — such as the smile you wear when you imagine you're in a room with someone you like.

As well as isolated academic jewels such as the Collegium, Hungary also boasts a healthier high-tech industry than many of its East European neighbours. Multinational companies such as Ericsson, Nokia and Hewlett-Packard have established research centres in Hungary. And the nation's drug industry, which used to provide some two-thirds of the Eastern bloc's needs, has survived the collapse of communism. Although companies such as Richter Gedeon, EGIS and Chinoin are minnows in an industry dominated by huge multinationals, they each spend about US\$20 million per year on research, and pay young PhDs roughly US\$1,000 per month.

That's more than the average salary of a university professor — and therein lies the problem for Hungarian science. With spending on research languishing at about 1% of the country's modest gross domestic product, many of the best minds in Hungary are



Making connections: Hungarian scientists believe that Budapest is ideally placed to act as a linchpin for science in the former Eastern bloc.

forced to migrate to the West, or to leave science altogether. “The main bottleneck is the poor technical infrastructure at universities and academy institutes,” says Balázs Gulyás, a neuroscientist who now works at the Karolinska Institute in Stockholm, Sweden.

Like many Hungarian researchers abroad, Gulyás maintains close ties with his home country. Some eventually do return. Zoltán Nusser, a group leader at the Institute of Experimental Medicine in Budapest, spent eight years working in Britain and the United States. Although he is pleased to be back home, Nusser can only pursue his research thanks to grants from Britain’s Wellcome Trust and the US Howard Hughes Medical Institute. These funds allow him to hire talented postdocs and buy expensive ‘patch clamp’ devices for electrophysiological recording.

All of this would be unthinkable if Nusser, like the majority of young scientists in Hungary, had to depend on domestic money. “There are many good scientists here, but only one out of a thousand can afford good equipment,” he says. The average size of grants provided by the Hungarian Scientific Research Fund, known by its Hungarian abbreviation OTKA, is just US\$10,000 per year — “far from enough”, admits its president, neuroendocrinologist Gábor Makara.

OTKA’s most prestigious funding programme, nominally intended for research leaders who will establish a ‘school’ of like-minded scientists, provides about 20 grants

per year of around US\$110,000 each. In theory, such grants might allow the likes of Gulyás to return home, or end the reliance of scientists such as Nusser on foreign funding. But the way in which the money has been distributed illustrates perhaps the biggest obstacle facing Hungarian science. The recipients have averaged 66 years of age, and are selected using backwards-looking criteria such as the number of PhD students they have produced over their careers — rather than their potential to make a real difference in the future.

Hampered by history

To the younger generation, this is indicative of the grip that the old guard, and its Soviet-era thinking, still has on the Hungarian scientific establishment. “Old asses who grew up in a contra-selective system run the institutes and referee each other’s grants — it’s a disgrace,” complains one young group leader. “This is what makes it really difficult here for talented young people to get on.”

Unlike its counterparts in neighbouring countries, the Hungarian Academy of Sciences, which runs some 35 institutes across the country, was never radically cleansed of its communist heritage. Leaders of the country’s universities have also resisted attempts to introduce an environment in which energetic young researchers can become independent and flourish.

The Academy of Sciences is at least now making positive statements about reform. “Yes, we’ll need to increase competitiveness,

give grants solely according to quality, and use more foreign reviewers,” says Vizi. But the body’s critics want to see action, and the harshest among them question whether the 70-year-old Vizi, despite a respectable international reputation for research, is the right person to catalyse such fundamental change.

Despite all the difficulties, the good news is that the pipeline of young Hungarian scientific talent shows no sign of drying up. Livia Meszaros, a medical student at Semmelweis University in Budapest, who in 2002 won the Hungarian Association for Innovation’s prize for the best high-school researcher in Hungary, argues that science still has a certain glamour: “We don’t have famous soccer players any more, and we never had big pop stars — so what else can you do than shine in science?”

The problem is that those who do shine are likely to join the brain drain to the West. Advocates for reform have grown cynical about the Hungarian scientific establishment’s willingness to create an environment that could begin to reverse this exodus. But perhaps they might be shamed into action by the presence of Western scientists brought into Hungary under EU programmes. “What it would really take to change things are foreigners, who are unhappy here, to complain and make a noise,” says Kondor.

Quirin Schiermeier is *Nature’s* German correspondent.

Collegium Budapest ♦ www.colbud.hu

Hungarian Scientific Research Fund ♦ www.otka.hu

Hungarian Academy of Sciences ♦ www.mta.hu

Automatic archaeology

Digging in the dirt in search of clues to the past has churned up millions of pieces of pottery. Haim Watzman unearths the new technologies being developed to sift through them all.

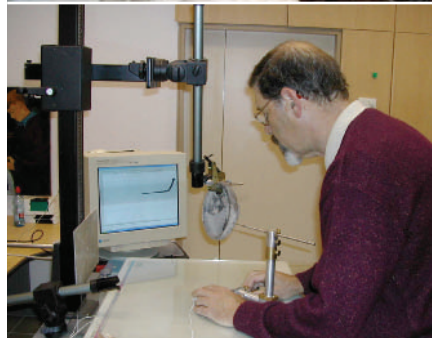
Vered Rosen is an artist. For the past few years, she has made her living drawing pieces of pottery using a simple sketchpad and a pencil, for a team of archaeologists excavating Tel Dor on the northern Mediterranean coast of Israel, south of Haifa. But now her employers are retraining her. They have sent her to the Weizmann Institute of Science in Rehovot to learn to draw using a profilograph, a device that can produce digital images of pottery fragments — images that can be analysed and classified by a computer.

For archaeologists, the ability to sift automatically through pottery samples is fast becoming a necessity. A typical excavation site might disgorge hundreds of thousands of pieces of broken clay pots — the favoured all-purpose vessels from ancient times until well into the Middle Ages — all of which need to be washed, sorted, labelled and analysed. The pots' shapes, dimensions, texture and style can help to date the layers of soil through which the archaeologists are digging, and can even reveal when and where different societies were trading with each other. Pottery shards are, to an archaeologist, as important as microfossils are to a palaeoclimatologist — the crucial clues that are used both to date a stratum and to understand the conditions that prevailed at the time. But there are not enough trained people to deal with the flood of pottery finds, and there are doubts in the community about whether the conventional method of hand-drawing is the best way to study them.

Artistic difference

Rosen's metamorphosis from artist to technician began when Uzy Smilansky — a physicist at the Weizmann Institute and an archaeological hobbyist — celebrated his sixtieth birthday three years ago. On that day he took his entire family to an excavation at Ein Gedi, on the northwest coast of the Dead Sea. "At the time I was looking for something new to do in the last phase of my career," says Smilansky. His experience as an amateur archaeologist meant that he was aware of the problem of comparing and analysing vast quantities of pottery. And his excavation holiday drove that home. "I thought it was time to harness the computer to the profession of archaeology," he says.

With that goal in mind, Smilansky was directed to Ilan Sharon of the Hebrew University of Jerusalem. Sharon, one of the chief



Helping hand: Uzy Smilansky (left) aims to use a drawing device called a profilograph to record the wealth of pottery shards (right) unearthed at sites such as Tel Dor (above and above right).

shards. But measurements for each piece had to be painstakingly taken and entered into the computer by hand.

Drawing on technology

Smilansky's team focused its efforts on eliminating that problem with the help of a profilograph. The device, invented years ago for use by architects, designers, artists and scientists, resembles a cross between a drafting table, a caliper and a computer mouse. To use it, a piece of pottery is clamped above the table in a position such that the central axis of the pot, if it were whole, would run parallel to the table. Then a delicate metal pointer attached to a computer mouse is used to trace the outline of the pot, clicking every few millimetres. This forms a series of data points in three-dimensional space that the computer then connects to create an image of the shard's surface.

Rosen is still learning how to use the device. At this point, she says, it takes her about an hour to trace a single piece — much longer than the 20 minutes it would take to do a simple pencil drawing of the fragment. But the end result is much more accurate and objective than any freehand drawing.

excavators at Tel Dor, has a similarly mixed background, having done two years of undergraduate work in mathematics before switching to archaeology. Sharon was enthusiastic about Smilansky's idea, and knew of a Hebrew University undergraduate named Avshalom Karasik who was doing a double major in archaeology and mathematics. Together with Ayelet Gilboa of the University of Haifa — the Tel Dor excavation's pottery expert — they made a perfect team.

Together, they soon learned that others have been in hot pursuit of new ways to study pottery in bulk. Statistician Clive Orton of University College London was one of the first to outline the potential benefits of speed and objectivity that an automated system could offer¹. In the late 1980s, Orton began to work on one of the first computer programs that can analyse and compare information about large quantities of



ZEEV RADOVAN/TEL DOR EXPEDITION

Smilansky's role, meanwhile, has been to program the computer to interpret these drawings. His program creates a 'curvature function' from the lines — a mathematical representation of features such as the shape and size of the pot. For the moment, his program emphasizes the shape of the top few centimetres and rim of the pot — a diagnostic characteristic that many archaeologists think is the best indicator of when and by whom the pot was made.

The technique's great advantage is its objectivity. "Hand drawings are inevitably biased, no matter how objective they try to be," says Gilboa. "The artist will often have the excavator's theories in mind when looking at a piece, and when the archaeologist reviews the drawing he'll often tell the artist to emphasize a feature that he thinks is important."

"Most drawings are not accurate," agrees

Robert Sablatnig, a computer scientist at the Vienna University of Technology in Austria. "Archaeologists somehow classify their findings in their mind and then draw what they think they should draw to make the classification correct. When you examine the original fragments, they often don't have the same features as in the drawing."

To test the usefulness of the profilograph, Gilboa, Karasik, Sharon and Smilansky turned their attention to an ongoing dispute about evidence of trade between Hazor, a site north of the Sea of Galilee that was a major urban centre during the Iron Age, and Tyre, a city in modern-day Lebanon that was once a major Phoenician port. A large bulk of evidence points to the fact that these two nearby cities were engaged in trade. But there have been arguments about whether there is additional evidence for this in the large collections of 'torpedo' jars found at both sites. These

long, cylindrical vessels, used in the northern Kingdom of Israel and Phoenicia during the Iron Age IIB period, about 2,800 years ago, were made to carry anything from grain to wine in commercial trade.

One of the Hazor researchers argues that all of these jars were manufactured in Hazor, and that some were exported to Tyre². But the archaeologist who found the jars in Tyre claims that, if anything, they went in the opposite direction — the Hazor jars were imported from Tyre³. A third study, by Gilboa, argued that the existence of tiny, but consistent, morphological differences between the Tyre and Hazor jars indicates that the jars are not the same at all, but were each made in their own home city⁴. Tests using the profilograph backed up Gilboa's contention — the jars from each city are far more similar to each other than to those from the neighbouring area. The discovery might not help to clear up any confusion about trade between the two cities, but it does show that archaeologists can benefit from a standard, objective system to judge similarities and differences in pottery.

Subjective view

But even Smilansky's technique cannot remove all subjectivity from the process — someone had to decree that rim shape is among the best deterministic factors, for example. "We began with a great deal of respect for archaeologists," Sharon says. "Our assumption is that the computer's job is to provide an emulation of the archaeologist's intuitive processes — with a grain of salt."

Nor will something as simple as an equation describing the rim of a pot ever completely replace the watchful eye of a trained archaeologist. "There's a lot more to identifying pottery than looking at drawings," says Jodi Magness of the University of North Carolina at Chapel Hill, an expert on Roman pottery. "It also involves familiarity with the fabric, which means things like colour and texture. You have to handle a lot of pottery to become familiar with it. Some pottery experts even taste pot shards to identify them."

"There's something about the human mind's ability to deal with data that so far computers don't do," says Aren Maier of Israel's Bar-Ilan University, who excavates at Tell es-Safi, a site southwest of Jerusalem. Maier is so far sceptical of the value of the simple drawings produced by computer programs such as Smilansky's. But, he adds, "it could be that that's enough — especially given our need to save time, space and resources".

That is exactly what the profilograph technique promises to do. "What do you do when you have 100,000 pot shards?" asks Gilboa. There are simply not enough artists to cover that much work. "There's no question that there's a ton of pottery and not



Given the wealth of findings from Tel Dor (above and right), a computer log of pottery shape will be a huge bonus. Future databases will also aim to incorporate information on clay type and colour.

enough people who can work on it," says Maier. Part of the problem is that drawing and comparing bits of pottery is seen as an unglamorous job. "It's the kind of work you generally do for a doctorate and then never touch again," says Gilboa. "The general feeling in the field — it's totally mistaken, of course — is that the study of pottery is for wusses," adds Maier.

But that means that masses of pottery from a site such as Tel Dor are never drawn or analysed, potentially leaving interesting facts undiscovered.

One reason that Tel Dor was chosen as a laboratory for automatic analysis is that the pottery in the area changed relatively rapidly over time, with fashions in the city shifting every few decades. During Israel's iron age, around 3,200–2,600 years ago, the profiles of pots shifted from looking like bent femur bones — with knobs and indentations adorning the top — to simpler S-shaped profiles resembling the horn of a longhorn cow, passing through myriad styles in between.

An archaeologist who chooses to publish drawings and studies of only one or even a handful of these pieces will probably miss something that could provide excavators at another site with crucial information about relations between the two communities, says Gilboa. Even if all the pottery is drawn and analysed by hand, it is likely that few people will bother to look them up in the weighty excavation reports — a process so slow and laborious that it is seldom carried out, she adds.

An automatic technique, together with a computerized database of the results, would solve these problems. But the profilograph is not the only possible technological solution. Sablatnig, for example, is working on a different experimental system that uses photographs to form a three-dimensional

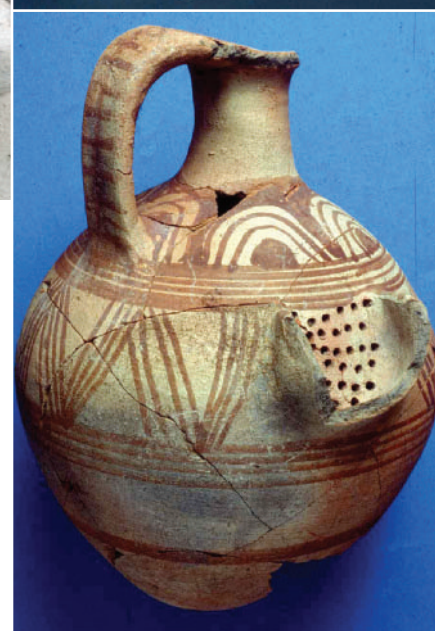
computer image of a broken pot. This process is faster than using a profilograph, so Sablatnig does not restrict himself to looking at the tops of pots. As a result, he gains a more complete picture of the reconstructed piece. But so far it is unclear which method will prove more powerful in comparative studies. Smilansky points out that his own curvature functions are more versatile than Sablatnig's, as archaeologists can use them to specify the aspect of a pot's shape in which they are most interested. The pair plan to collaborate to devise the best possible method — one option that Smilansky is already investigating is the possibility of using lasers to image the pieces.

Shaping up

Others, including research teams at Brown University in Providence, Rhode Island, and Arizona State University in Tempe, are also tackling the problem.

But many of these systems suffer from the fact that they address only one element of the pottery — its shape. What's more, to reconstruct the image of a pot from a few shards, they assume that the vessels they are looking at are basically symmetrical. Prudence Rice, an archaeologist at Southern Illinois University in Carbondale, points out that people from the Americas did not use potter's wheels, so this assumption does not apply to pieces from the New World.

Instead, archaeologists rely more heavily on other criteria, such as colour or the composition of the clay, to identify such pieces. "I don't see how these three-dimensional images are going to replace that," says Rice. Smilansky and Sablatnig are working together to develop imaging methods that can capture features such as painted pictures on the pots. And Orton, one of the pioneers of automated comparative archaeology, is working



ZEEV RADOVAN/TEL DOR EXPEDITION

on adding information about colour, texture and clay type into his databases of pottery shards. But he has not been able to automate the collection of that kind of data — he is still doing it all by hand.

In the end, even if these teams succeed in establishing profilographs or laser-imagers as standard equipment on archaeological digs, the kind of comparisons that Gilboa wants to carry out will also require backlogging information from previously published excavations into a computer format — an immense amount of work that Rosen and her colleagues could never achieve. Gilboa nevertheless dreams of an international and comprehensive pottery database that will make serious comparative work not only possible, but a pleasure. Maybe then, she thinks, archaeologists can begin to answer questions that they have not yet been able to address, such as whether pot shape changed over time because of fashion trends, new technologies or both. Who knows? Maybe working on pottery will even become glamorous again. ■

Haim Watzman is a freelance writer in Jerusalem.

1. Orton, C. *Sampling in Archaeology* (Cambridge Univ. Press, Cambridge, 2000).
2. Geva, S. *Bull. Am. Sch. Oriental Res.* **248**, 69–72 (1984).
3. Bikai, P. M. *Bull. Am. Sch. Oriental Res.* **258**, 71–72 (1985).
4. Gilboa, A. *Qedem Rep.* **2**, 15 (1995).

Cooperation needed to increase fertilizer efficiency

Better use of nitrogen could provide more food and reduce the environmental impact.

Sir—Your News Feature “Fertilized to death” (*Nature* 425, 894–895; 2003) deals with an extremely important issue: reactive nitrogen and the need to increase the efficiency of nitrogen fertilizers. James Galloway’s work, as well as that of the IVL Swedish Environmental Research Institute, has pointed to significant environmental problems with excess reactive nitrogen.

Although environmental concerns about excess nitrogen are worth highlighting, the need to produce more food, more efficiently, in many parts of the world is even more urgent. Energy expert Vaclav

Smil estimates that approximately 40% of the world’s dietary protein supply in the mid-1990s originated from fertilizer nitrogen (V. Smil *Enriching the Earth* 156–161, MIT Press; 2001). Agronomists and soil scientists throughout the world are working to make nitrogen use in farming more efficient for both food production and environmental reasons.

Although cutting back on fertilizer use might be a method to reduce the total load of reactive nitrogen in developed countries with excess food production capacity, it is not an acceptable solution in countries with malnourished people. The effect of

increased food costs on people with low incomes in developed countries must also be considered. Efficiency improvements can help reduce nitrogen-fertilizer consumption if populations stabilize, and reduce its impact on the environment.

Ecologists, agronomists, environmental groups and industry must join to increase nitrogen-fertilizer efficiency for the benefit of everyone. Publications such as *Nature* should lead the way in building interdisciplinary support for this worthy goal.

M. M. Alley

Crop and Soil Environmental Sciences Department, Virginia Tech, Blacksburg, Virginia 24061-0403, USA

Fertilizer: complex issue calls for informed debate

Sir—The News Feature “Fertilized to death” (*Nature* 425, 894–895; 2003) contains several inaccuracies about how reactive nitrogen is affecting ecosystems. As scientists working to understand these effects and to formulate methods for reducing them, we feel that the increase in reactive nitrogen is a serious and critical issue, which deserves informed discussion of the many complex issues involved.

First, although excess fertilizer use can have adverse environmental consequences, it contributes little to acid rain, whose principal causes are sulphur and NO_x emissions from fossil-fuel combustion by power plants and vehicles. Moreover, although ammonia emissions from agriculture contribute to high rates of nitrogen deposition in some forests, indirectly increasing soil acidity, these contributions are primarily from livestock, not from fertilizer.

Second, although the feature does state that it is not clear how different types of forest will respond to increased nitrogen deposition, it should be emphasized more clearly that the effects of increased nitrogen deposition on forest growth and mortality are highly variable and complex, and even the direction of response (increased or decreased growth) remains uncertain.

Third, it is important to recognize that almost half the nitrogen applied to crop fields is lost through aqueous run-off and atmospheric emissions (not run-off alone as stated in the article) and that the environmental impact of these two processes are very different. Aqueous losses of fertilizer nitrogen (~20%) are usually substantially less than atmospheric losses, although this run-off can still have serious environmental effects. Much of the loss to

the atmosphere is through denitrification, which removes most of the reactive nitrogen from the ecosystem but can produce the greenhouse gas N₂O.

Finally, not all potential solutions to this problem need to be expensive or result in increasing the costs of agricultural production. In the United States, at least, current levels of agricultural production can be met by reducing excess application of synthetic nitrogen fertilizer, potentially saving farmers money.

Mark B. David*, Gregory F. McIsaac*, Robert W. Howarth†, Christine L. Goodale‡, Laurie E. Drinkwater‡

**Department of Natural Resources and Environmental Sciences, University of Illinois, Urbana, Illinois 61801, USA*

†Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853, USA

‡Department of Horticulture, Cornell University, Ithaca, New York 14853, USA

Fertilizer: no-till farming could reduce run-off

Sir—In the News Feature “Fertilized to death” (*Nature* 425, 894–895; 2003), your reporter repeats a claim sometimes made in the medical literature that blue-baby syndrome (methaemoglobinemia) is caused by excess nitrate in the tap water used to make up infant formula. However, it is nitrite, synthesized from nitrate by bacteria, that is the ultimate culprit in methaemoglobinemia.

As humans cannot convert nitrate present in food or water to nitrite by themselves, the most likely source of bacterial contamination in cases of blue-baby syndrome is poor hygiene in making up infant formula. Fortunately, infants acquire an enzyme by the age of six months

that protects them against nitrite poisoning.

It is true, of course, that excess nitrate levels in tap water are caused by fertilizer run-off and leaching into rivers, but it is surprising that there are far fewer cases of blue-baby syndrome in Europe (which has much higher levels of nitrate in tap water) than in the United States.

It is also worth noting that, despite hopes that organic farming would reduce nitrate in tap water, recent studies show that nitrate leaching from well-managed organic and conventional farms is effectively identical^{1,2}.

A less-advertised, and cheaper, way of reducing nitrate levels in drinking water is ‘no-till’ agriculture. The plough has been used as a means of controlling weeds since the advent of agriculture itself. However, methods whereby crop material is mulched and left on the surface, and weeds controlled by herbicides, have been found to produce after several years a soil structure that is highly resistant to erosion, run-off and leaching.

Nitrate loss from a fertilized no-till field is approximately 12–20% of that from the equivalent ploughed field. Greater use of no-till, which is increasingly common in the United States but receives little encouragement from governments in Europe, would probably see nitrate levels in tap water return to those found before 1950 — about a quarter of current levels.

Anthony Trewavas

Institute of Cell and Molecular Biology, University of Edinburgh, King’s Buildings, Edinburgh EH9 3JH, UK

1. Pickett, J. & Goulding, K. House of Lords 16th report *Organic Farming and the European Union* 166 (HMSO, London, 1999).
2. Stopes, C. et al. *Soil Use Management* 18, 256–264 (2002).

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter.

When development goes wrong

Studying mutants helps our understanding of how we all take shape.

Mutants

by Armand Marie Leroi
 Viking, 2003. 320 pp. \$25.95

Peter Little

Nature and nurture have a capricious power to alter body form, often with monstrous outcomes. Only relatively recently has it become morally unacceptable to collect and exhibit human examples. At the same time, huge strides have been made in understanding the molecular events that pattern and define the form of living creatures. It is perhaps ironic that this understanding has often been based on the deliberate creation of equally deformed model organisms.

Armand Leroi has realized that our expanding knowledge of the molecular underpinnings of development means that we may now be able to revisit these unfortunate humans, no longer to gape and stare, but to see if we can apply our understanding to the only organism in which we are really interested — ourselves. The result of this realization is his new book, *Mutants*.

Leroi has an extraordinarily extensive familiarity with a dazzling range of information that stretches from Neolithic sculpture to molecular developmental biology in the pages of *Nature* in 2002, and he draws on this to detail the remarkable molecular catastrophes that have been visited upon some unfortunate humans. He seeks to relate these cases back to a modern explanation based on molecular developmental biology, and in so doing to illustrate the principles that lie behind our development.

The story unrolls in three critical chapters that cover the control of overall body form. Conjoined twins lead us through the establishment of the three basic tissue types and the power of the 'organizer' to direct their development. Next, the one-eyed Cyclops of Greek mythology and terrible images of its human counterparts lead us to the wonderful elegance and simplicity of the homeotic genes and their ability to specify the identity of body parts. A chapter on limb development and failures to control limb induction then rounds off this most fundamental level of control of body form.

So far, *Mutants* might sound like yet another textbook, but this could not be further from the truth. What textbook would start a section on limb development with the story of two religious dissenters, Margaret McLaughlin and Margaret Wilson, who were tied to a stake in the River Bladnoch in Scotland and left to be overwhelmed by the rising tide and the estuarine crabs? And what's it got to do with development? You may well



Altered morphology: Lavinia Fontana's *Portrait of a Girl Covered in Hair*.

wonder. Their executioner was supposedly 'punished' for his cruelty by his children being born with hands like crabs' claws. Leroi uses this story to tease out the controls of limb development through tales of crook-footed Hephaestus, the two-toed Wadoma tribe of Africa, the armless fiddler of Germany, the handless and footless aleijadinhos of Brazil, Le Petite Pepin with his thalidomide-like shortened limbs, and finally even Charles Darwin (for his views on polydactyly). Leroi uses this wild and exotic cast of characters as the basis for discussion.

The next four chapters gallop through the control of skeleton development (or lack of it, in the case of Harry Eastlake, who died at 40, his normal skeleton encased in another); the control of size (here we find pygmies, giants, monsters and castrato opera singers);

gender (illustrated by the amazing observation that hyenas give birth through an enlarged clitoris, and by the tragic suicide of Abel Barbin, born as a girl to die as a man); and, finally, skin and hair, where we meet albinos, piebalds and the hairy Shwe-Mong and his children at home in the Burmese court.

At this point, with only the briefest pause for breath, Leroi charges into the penultimate chapter, on ageing. This is a departure from the more morphological focus of the work so far and, for me, is less exciting. Ageing has all the features of being an innate property of an organism's biological constraints, and I confess to being prejudiced — it is too hard for us right now and is better left alone.

The final chapter, though, is wonderful.

We all choose, or are driven, to be excited not by all of science itself, but by some more limited domain; Leroi is driven by a fascination with the morphology of race and, more personally still, by the biology of beauty. I too know the effect of a mutation in the MC1R gene, which is responsible for redheads, and I know that it is beautiful. Leroi draws tight his net of wonderful human diversity and gracefully displays its contents, and I am full of admiration for his willingness to build on his declared prejudice.

Despite its subject matter, *Mutants* is an exquisitely life-enhancing book. It captures what we know of the development of what makes us human, and it recognizes the random tragedy inflicted by nature and nurture. Read the book and you will be exposed to both a scientific world that no longer exists and to that of the twenty-first century. Read it and you will know a tiny part of what it is that has made you the person you are. Read it and enjoy words written carefully, elegantly and with sensibility. Read it and marvel. ■

Peter Little is at the School of Biotechnology and Biomolecular Science, University of New South Wales, Sydney, New South Wales 2052, Australia.

Watching the Earth move

Pendulum: Léon Foucault and the Triumph of Science

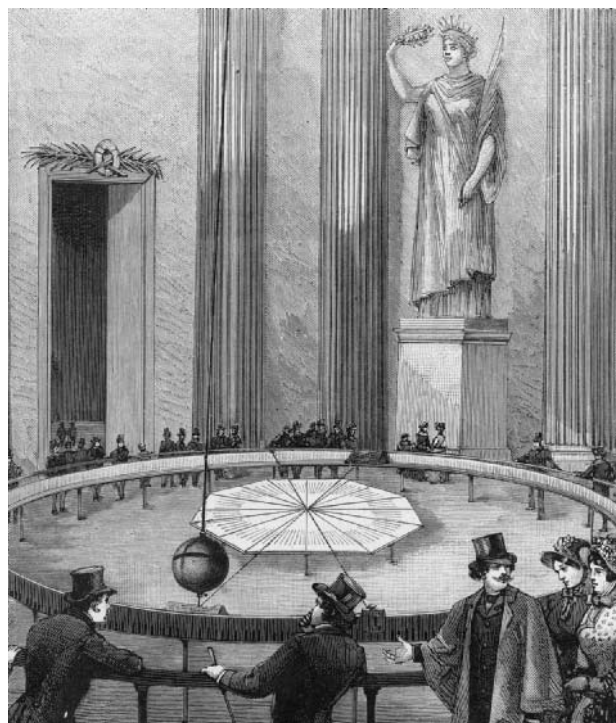
by Amir D. Aczel

Atria: 2003. 288 pp. \$25

Michael Matthews

Léon Foucault (1819–1868) is known to millions by virtue of science museums around the world displaying his pendulum. Foucault's pendulum gained even wider recognition as the title of a best-selling novel by Umberto Eco. But although Foucault's name is widely recognized, knowledge about him, even among scientists and historians of science, is thin. This scientific and social biography by Amir Aczel, who has written several popular-science books including *Fermat's Last Theorem* (Four Walls Eight Windows, 1996), should rectify this problem.

Foucault experienced similar travails to those endured by John Harrison, another pendulum enthusiast and the protagonist of Dava Sobel's best seller *Longitude* (Walker, 1995). Both men came from outside their country's scientific establishment, both were long rejected by the establishment, both were given recognition by 'royal' decree, and both enjoyed their rightful fame only at the end of their lives. But whereas Harrison was predominantly a technician who worked for 17 years on the construction of a single pendulum clock, Foucault had



In full swing: Léon Foucault's pendulum revealed the Earth's rotation.

much wider technical and scientific interests.

As a small boy, Foucault suffered numerous illnesses. One of his eyes was myopic and the other was far-sighted. A friend wrote of him that "the feebleness of his constitution and the slowness that characterized his work made it impossible for him to frequent a college". He could therefore pursue his studies only with help from a devoted tutor, "under the watchful eyes of his mother". But he did have a gift for working with his hands.

The combination of dogged and skilled handiwork with scientific interest and competence were the foundations of his scientific successes. For example, he improved the daguerreotype method to obtain the first ever 'photograph' of the Sun showing its sunspots. He also constructed a massive (190-centimetre) but finely accurate reflecting telescope that was engine-driven to compensate for Earth's rotation; this instrument contributed significantly to Urbain-Jean-Joseph Le Verrier's astronomical triumphs. In 1850 Foucault determined the speed of light to be 298,000 kilometres per second, which is within 0.6% of the current accepted value. And in 1851 he made the world's first gyroscope that could be pointed towards true north, thereafter holding its position and thus being a non-magnetic and superbly accurate compass.

But Foucault's fame rests on his pendulum demonstration, which rightly takes centre stage in Aczel's book. On 2 February 1851, Foucault sent to all known scientists in Paris a simple invitation that read: "You are invited to come to see the Earth turn, tomorrow, from three to five, at the Meridan Hall of the Paris Observatory." The following day,

scores of scientists, mathematicians and savants duly gathered behind a wooden balustrade that marked off the oscillation space of an 11-metre heavy pendulum that swung along the marked-out Paris meridian. Once the pendulum was set swinging, the Earth moved before the eyes of the audience. The plane of oscillation of the pendulum slowly but perceptibly altered. Foucault's clever handiwork had ensured that there was no torque force on the pendulum, so the only conclusion to be drawn was that the Earth had moved.

This was the first dynamical proof of Earth's rotation — something that had eluded natural philosophers and scientists since Pythagoras and

Aristarchus in the ancient world proposed that it was Earth that rotated, rather than the heavens that rotated around a stationary Earth. Copernicus had resurrected the heliocentric, rotating-Earth theory in 1543; Galileo defended the view in the early seventeenth century; Newton and other leading astronomers and scientists also adopted the position. But with enough determination and conviction, all of their astronomical evidence could be reconciled with the stationary-Earth theory. Tycho Brahe took this option, and so did the Catholic Church, which kept all 'rotating Earth' publications on the Index of Forbidden Books until as late as 1835.

Aczel's book benefits from his study of original sources, including those in the French Academy of Sciences, the French National Library and the Paris Observatory. The book nicely connects Foucault to his political, social and scientific times. It details the generous support that Foucault received from Louis-Napoléon Bonaparte after the scientific establishment had all but ignored him. Aczel rightly highlights the question of why it was that so simple a demonstration of the moving Earth had eluded the roll-call of geniuses who had sought it. And he also asks why the French Academy of Sciences resolutely refused membership to Foucault not only before his demonstration, but for 15 years afterwards. ■

Michael Matthews is at the School of Education, University of New South Wales, Sydney, Australia. He is the author of *Time for Science Education: How Teaching the History and Philosophy of Pendulum Motion Can Contribute to Science Literacy* (Kluwer, 2000).

A glimpse inside the Sun

Sunquakes: Probing the Interior of the Sun

by Jack B. Zirker

Johns Hopkins University Press: 2003.

288 pp. \$29.95, £22

Jørgen Christensen-Dalsgaard

"What appliance can pierce through the outer layers of a star and test conditions within?" asked Arthur Eddington in his book *The Internal Constitution of the Stars* (Cambridge University Press, 1926). In the case of the Sun, this question has been answered dramatically in the past few decades by the emergence of helioseismology — and there are high hopes that such investigations could be extended to other stars.

The surface of the Sun shows a variety of dynamic phenomena, largely associated with convection, where hot plasma rises to the surface. These dominate energy transport in the convection zone, which occupies the outer part of the Sun. It is the convective cells that give the Sun's surface its granular appearance. These and other motions are also reflected in the velocity of the solar atmosphere, as measured through the Doppler effect.

Variations in the Doppler velocity at the solar surface may have been noted in 1916 by Harry Plaskett, and were unambiguously detected by A. B. Hart in about 1955. In 1961, Robert Leighton, John Evans and colleagues revealed that this signal seemed to be oscillating with a period of about five minutes. The physical nature of these oscillations remained uncertain for more than a decade, until in 1970 Roger Ulrich, and independently John Leibacher and Robert Stein, proposed that the oscillations arise from trapped acoustic modes resonating in the solar interior. This was confirmed by Franz-Ludwig Deubner in 1975 and, a little later, by Edward Rhodes and his collaborators. With these results, and the detection in the late 1970s of five-minute oscillations propagating throughout the entire Sun, the potential for using observations of these surface oscillations to probe the solar interior was evident, and helioseismology was born.

In his rich and enjoyable book, Jack Zirker presents the history of helioseismology, from the work of Leighton and Evans to current detailed investigations of the Sun's internal structure and dynamics. His discussion of the early days quite rightly includes the apparent detection of oscillations in the

solar diameter by Henry Hill and his group, and the observation of an enigmatic 160-minute oscillation. Although these results were not confirmed by later investigations, they did help to inspire the subsequent developments. Descriptions of the investigations of flows in the solar convection zone associated with the solar magnetic cycle, and the peculiar periodic variations in rotation near the base of the convection zone, bring the book essentially up to date. Also covered is the development of asteroseismology, including the impressive results obtained from analysis of oscillations of white dwarfs, and the early results from emerging studies of solar-like oscillations in other stars.

A crucial feature of helioseismology is the way in which it can throw light on other aspects of the solar interior and the physical processes that control it. A key problem in astrophysics since 1968 has been the discrep-

aspects of solar magnetic activity. As discussed by Zirker, helioseismic determinations of solar internal rotation were not consistent with initial numerical models that had attempted to describe the dynamics of the solar convection zone and the origin of the variation of the solar surface rotation rate with latitude on the Sun. This led to a revision of the models of the generation of the magnetic field, which now seems likely to be dominated by the region of strong radial variation in the rotation rate at the base of the convection zone. Also, more sophisticated models of the interaction between convection and rotation, which better capture the turbulent nature of the flow, are beginning to yield results in accordance with the helioseismic inferences. And finally, local helioseismology has provided information about the subsurface structure of solar active regions and their emergence, and even led to monitoring of solar activity on the far side of the sun.

Zirker's excellent book combines the development of helio- and asteroseismology with a lucid and detailed overview of modern solar physics. The text is at a level that is accessible to the interested layperson, with more technical aspects discussed in a substantial set of notes. There are a few minor inconsistencies and misprints, typically in names, and it would have been useful to have included a brief bibliography for those readers interested in delving further into particular topics.

The book is wholeheartedly recommended. Older partici-

pants in the field will enjoy being reminded of how helioseismology has developed; younger scientists get an introduction to the history of their subject, and the general reader is given a fascinating overview of one of the newest and most active fields of astrophysics, including enjoyable cameos of some of the participants.

The appearance of this book is also timely; helioseismology is a well-consolidated field and asteroseismology is poised on the brink of full emergence. It is unfortunate that the European Space Agency's Eddington mission, which was to have carried out asteroseismic investigations for a large number of stars and searched for extra-solar planets, has been cancelled. Should it be reinstated, we would obtain a fuller answer to Eddington's original question, and our understanding of the interior properties of other stars and their place in the Universe would be greatly advanced. ■

Jørgen Christensen-Dalsgaard is at the Theoretical Astrophysics Center, Danish National Research Foundation, and the Institute of Physics and Astronomy, University of Aarhus, DK 8000 Aarhus C, Denmark.



A simulation of turbulent convection mirrors activity at the Sun's surface.

ancy between the predicted rate of production of neutrinos in the solar core and the rate of neutrino detections in several independent experiments. This initially raised suspicions about possible problems in solar modelling, and various changes to the models were proposed to bring the neutrino predictions into line with the observations. However, even early helioseismic results strongly indicated that these modified models could not be right, and that therefore some other explanation was needed for the measured neutrino flux. Zirker gives a very clear presentation of this problem and its final solution: the Sudbury Neutrino Observatory in Canada recently demonstrated that the so-called electron neutrinos generated by nuclear reactions in the solar core undergo partial transformation to other neutrino types en route to the Earth, the combined neutrino flux being in full agreement with the models.

Rotation and other flows in the solar interior are closely related to the generation of the solar magnetic field and the 11-year cycle in the number of sunspots and other

A glass bead game

Thomas M. Bayerl

Microscopic glass beads coated with lipid membranes provide a sensitive detector of interactions between proteins and ligands. The changing spatial order of the array of beads in solution is the key.

At the start of a game of snooker, the balls are arranged in an ordered shape, each touching its neighbours, in a triangular rack. As the game begins, a player aims the cue ball at the apex of the triangle and disperses the object balls over the table. The sudden change in order, triggered by the transfer of momentum from the cue ball, is a kind of (irreversible) 'phase transition', from a condensed, ordered state to a dispersed, disordered one. No one would imagine that the object balls could take on their initial triangular arrangement spontaneously, just by being dropped all at once onto the billiard cloth. Nor would they think that the phase transition could be achieved without transferring momentum from the cue ball. But, in their paper on page 139 of this issue, Baksh *et al.*¹ suggest that, with billiard balls shrunk to the size of microscopic beads and the billiard cloth replaced by a smooth glass or plastic surface, this is exactly what can happen.

The spontaneous self-assembly of a disordered bunch of microscopic glass (silica) beads dispersed in water into highly ordered, two-dimensional colloidal crystals — akin to preparing a snooker game without needing the triangular rack — is already the subject of intense research. Initiating a phase transition in the colloid without adding external momentum (the cue ball) or thermal energy is, however, a new achievement. Yet the only additional ingredients are simple: the surfaces of the beads are coated with an ultra-thin organic film, and molecules designed to interact with this film are dissolved in the surrounding aqueous solution.

Through a light microscope, Baksh *et al.*¹ saw that silica beads, 5 µm in diameter, each encased in a shell of 5-nm-thick phospholipid bilayer, form two-dimensional colloidal crystals spontaneously. The degree of order in the colloidal crystals depended on whether or not ligands in the bilayer could bind to proteins (receptors) dispersed in the surrounding solution. The choice of a phospholipid bilayer as the bead coating is not accidental. Lipid bilayers are the basis of all biological membranes. When applied as a coating to glass beads, these so-called supported bilayers have a distinct advantage, compared with covalently bead-bound molecules or self-assembled monolayers — they retain the fluidity and functionality of a natural membrane and can accommodate

ligands or receptors without loss of their functional activity. Since the first supported bilayer was fabricated² in 1985, they have been studied extensively, and have been applied in biosensors³, in advanced bioseparation techniques⁴ and in high-throughput molecular screening⁵.

A series of time-sequence images taken by Baksh *et al.*¹ (Fig. 3a on page 140) reveals that the colloidal crystals can be disrupted, just like billiard balls: there is a phase transition from a condensed to a dispersed state when receptor protein is added (Fig. 1). The forces involved are essentially van der Waals attractions between the beads, and the attractive and repulsive electrostatic forces that arise from both the bead surface and the bilayer. The binding of the receptor to the ligand at the bilayer surface modulates the 'pair interaction potential' between adjacent beads, thereby disrupting the condensed phase in a sudden transition. As a result, binding of the receptor can be quantified in a straightforward way, by calculating a 'pair distribution function', which encapsulates how close and ordered the beads are in each state. By adding different proteins, and even complex protein mixtures, to the solution, Baksh *et al.* proved that the transition from ordered to dispersed state is triggered only if specific binding between the bilayer-attached ligand and the dissolved receptor occurs. In contrast, the presence of non-specific, or even competitive, binders did not result in a phase transition — but neither

did it diminish the response of the system to the specific receptor.

The science of this technique is intriguing; using it to detect specific protein-binding events, without the need to use labels such as fluorescent proteins, is an interesting prospect. The method works at very low concentrations (in the pico- to nanomolar range), and could even be automated for screening purposes. The collective nature of the phase transition makes it easy to watch through a microscope, but direct observation is not necessary. Instead, the transition can be detected by analysis of the dispersed phase, by calculating its pair distribution function. Automated screening could be carried out using microwell plates with transparent bottoms, into which ligand-coated beads could be pipetted in the presence of potential receptor proteins. After allowing the beads to settle, the pair distribution function could be read out from an automated imaging system.

In his novel *The Glass Bead Game*, Hermann Hesse never described explicitly how the game worked. Similarly, Baksh *et al.*¹ do not offer a comprehensive description of the molecular mechanisms underlying the phase transition in their own glass bead game. Neither do they elaborate on how different receptors might manifest themselves through distinguishable pair distribution functions. But, as the scholars in the novel devoted their lives to the endless ramifications of the glass bead game, this work by

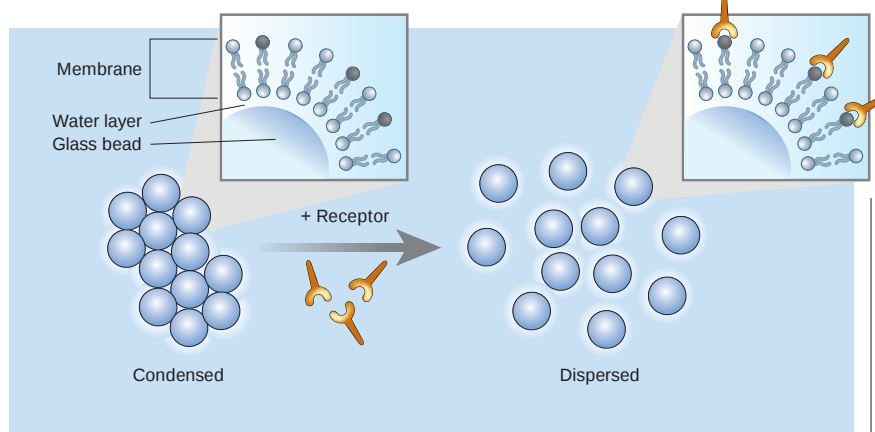


Figure 1 In an aqueous solution, microscopic glass beads, each coated with a lipid membrane, assemble into an ordered two-dimensional crystal. Baksh *et al.*¹ show how that order can be disrupted by adding to the solution a receptor protein that binds to a ligand at the membrane surface. This dramatic transition from an ordered to a dispersed state could be built into an automated scheme of detection for protein–ligand interactions.

Baksh *et al.* may open the door to the automated characterization of a wide range of complex molecular interactions that are at present poorly understood.

Thomas M. Bayerl is at the Physical Institute, University of Würzburg, Am Hubland, 97047 Würzburg, Germany.

e-mail: bayerl@physik.uni-wuerzburg.de

1. Baksh, M. M., Jaros, M. & Groves, J. T. *Nature* **427**, 139–141 (2004).
2. Tamm, L. & McConnell, H. M. *Biophys. J.* **47**, 105–113 (1985).
3. Sackmann, E. *Science* **271**, 43–48 (1996).
4. Loidl-Stahlhofen, A., Kaufmann, S., Braunschweig, T. & Bayerl, T. M. *Nature Biotechnol.* **14**, 999–1002 (1996).
5. Loidl-Stahlhofen, A. *et al.* *Adv. Mater.* **13**, 1–6 (2001).

Stem cells

How to make eggs and sperm

M. Azim Surani

Embryonic stem cells can develop into many specialized cell types in culture dishes. It now seems that they can also generate primordial germ cells, which then go on to form sperm and eggs.

A fertilized egg is potentially immortal: this fusion of egg and sperm gives rise not only to a new individual, but also (theoretically at least) to an endless series of generations. Three groups now suggest that it is possible to generate both of these remarkable cells — known collectively as germ cells — in a culture dish. Geijsen and colleagues¹, writing on page 148 of this issue, and Toyooka *et al.*², writing in *Proceedings of the National Academy of Sciences*, describe how they obtained sperm-like cells from mouse embryonic stem cells (ES cells) *in vitro*. Geijsen *et al.* even discovered that injecting their

sperm-like cells into natural mouse eggs resulted in early embryonic development. Meanwhile, Hübner *et al.*³ reported earlier this year in *Science* that they have succeeded in obtaining egg-like cells from mouse ES cells. So it is possible to produce germ cells with at least some attributes of sperm and eggs *in vitro*. These findings raise the possibility of deriving similar germ cells from human ES cells in culture — an idea that raises ethical issues as well as the prospect of unprecedented medical advances.

ES cells derived from five-day-old mouse or human embryos ('blastocysts') have the exceptional potential — depending on the culture conditions — to either multiply indefinitely or develop into an array of specialized cells⁴. To try to persuade mouse ES cells to generate eggs and sperm (Fig. 1), Geijsen *et al.*¹ and Toyooka *et al.*² allowed aggregates of the cells to differentiate into structures that somewhat resemble early embryos; Hübner *et al.*³ allowed ES-cell aggregates to undergo random differentiation spontaneously. The result was that, in the embryo-like structures and among the randomly differentiated cells, there were cells resembling primordial germ cells, which the authors detected by the expression of certain marker genes. The authors then isolated some of these primordial germ cells — the precursors of sperm and eggs — and allowed them to proliferate in culture. Curiously, all three groups noticed that the gene-expression pattern of the primordial germ cells showed highly accelerated development. Developmental timers are normally exquisitely regulated, so it will be important to know why the timing went awry in these cases.

The next stage involved transforming the primordial germ cells into sperm or eggs, which is a complex process that, *in vivo*, occurs within specific microenvironments^{5,6}. To achieve this *in vitro*, the groups adopted different approaches. Geijsen *et al.* allowed the process to occur spontaneously in the embryo-like structures; Toyooka *et al.* cultured the primordial germ cells with normal

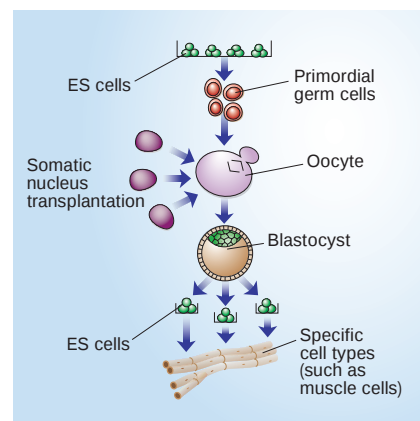


Figure 2 A possible use for eggs derived from ES cells — nuclear transplantation. As shown in Fig. 1, oocyte-like cells could be made in culture from ES cells, by way of primordial germ cells³. After stripping the oocytes of their own genetic material, they could be used as recipients for nuclei from adult (somatic) cells such as skin cells. The somatic nucleus could then be 'reprogrammed' by factors present in the oocyte, which is then allowed to develop to the blastocyst stage. Blastocysts contain epiblast cells, from which new ES cells can be derived. Each type of ES cell will inherit some properties of the adult donating the somatic nucleus, such as a propensity for certain complex diseases. The ES cells could then be used to derive specific cells with which to study the progression of those diseases, and perhaps to generate treatments¹⁰.

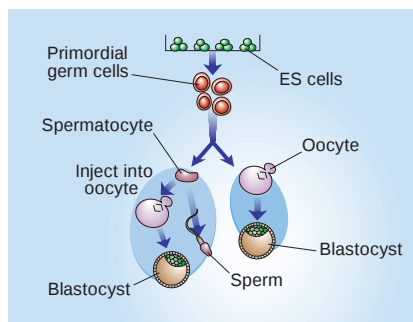


Figure 1 Germ cells from stem cells. In two of the new papers^{1,2}, embryonic stem cells (ES cells), cultured as aggregates, formed structures resembling early embryos ('embryoid bodies'); in a third paper³, the aggregates were permitted to undergo spontaneous differentiation. In all cases, a proportion of the ES cells produced primordial germ cells. Left, Geijsen *et al.*¹ found that some of these cells differentiated into spermatocytes (precursors of sperm) in the embryoid bodies. Moreover, injecting the spermatocytes into unfertilized eggs led to development to the early blastocyst stage. Toyooka *et al.*² found that culturing primordial germ cells with cells from fetal testis also produced sperm. Right, Hübner *et al.*³ isolated, re-aggregated and cultured primordial germ cells, which formed complex structures within which were found developing eggs. Release of these oocytes from the surrounding cells led to spontaneous activation and development to the blastocyst stage.

fetal gonadal cells. In the work by Hübner *et al.* the primordial germ cells were allowed to form aggregates again. Either sperm-like^{1,2} or egg-like³ cells were produced (possibly depending on the procedure used). At this time, the number of chromosomes must be halved to allow male and female germ cells to make equal genetic contributions at fertilization⁷. This apparently occurred, although additional confirmation would be desirable.

Developing sperm and eggs must also acquire their characteristic identity tags, or 'imprints'⁸, which regulate their complementary functions when embryonic development begins after fertilization. However, it is not yet known whether the eggs and sperm now generated^{1–3} have the appropriate imprints. This information will be crucial, because sperm and eggs can seem normal even without the appropriate marks — but their functions will be affected after fertilization when development starts⁸. So, although Geijsen and colleagues did obtain blastocysts when they injected their sperm-like cells into unfertilized eggs, we cannot make any predictions about the long-term development of these early embryos. It is also interesting that the eggs generated by Hübner *et al.* developed spontaneously to the blastocyst stage once released from the surrounding cells — even though mature eggs should remain 'arrested' until they are fertilized or

artificially activated. Encouragingly, these blastocysts exhibited an appropriate gene-expression pattern. But their full developmental potential is unknown.

These caveats aside, the three papers^{1–3} represent steps towards deriving primordial germ cells and, subsequently, eggs and sperm from ES cells. In the future, it will be important to develop means of controlling ES-cell differentiation into germ cells more tightly: in many respects, current methods rely on spontaneous and stochastic events, which makes it difficult to analyse each of the complex steps leading to the production of sperm and eggs. Systematic studies of germ-cell specification and properties (see, for example, refs 8, 9) will benefit from greater control over these steps *in vitro*.

So, what could be done with the 'synthetic' eggs and sperm? At present we are largely in the realms of 'fantastical thought experiments' — can we, for instance, generate viable embryos from synthetic germ cells? This could find applications in animal breeding, although researchers have yet to make ES cells from most mammalian species. Perhaps it might also prove possible to derive germ cells from human ES cells. If so, it would allow studies of human germ cells, about which very little is known. Use of such cells might also illuminate the causes of infertility and germ-cell tumours. It might even be possible to use synthetic sperm to treat male infertility. And, with improvements, the culture system could be used to examine many complex processes, including the roles of key genes and the mechanisms underlying imprinting and the halving of chromosome numbers.

Perhaps more importantly, however, a limitless supply of human eggs derived from existing ES-cell lines could have a radical impact on medicine. These synthetic eggs need not be perfect because, stripped of their own genetic material, they could be used as

recipients for nuclei or genetic material from adult cells such as skin cells. If such reconstituted eggs can 'reprogramme' the adult nucleus and develop to the blastocyst stage, researchers could derive new ES cells from them (Fig. 2). These cells in turn could be prompted to produce specific cell types for transplantation, to treat specific human conditions. Progress in this area is currently hampered by the scarcity of human eggs, the use of which also involves legitimate ethical considerations.

Human eggs derived in culture could also have an even more exciting use. By following the same procedure, it might be possible to use these eggs to generate ES cells that produce diseased tissues — the adult nuclei for the process being taken from patients with complex diseases such as diabetes. As noted at a meeting earlier this year, such ES cells would provide an unlimited resource, allowing approaches to the study of disease that are currently impossible¹⁰. This might, in turn, lead to new treatments.

And simply being able to study human germ cells in culture might allow more thorough investigations into the origin and properties of these remarkable cells. This could give us a grip on our destiny in more ways than we can imagine. ■

M. Azim Surani is at the Wellcome Trust Cancer Research UK Gurdon Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK.
e-mail: as10021@mole.bio.cam.ac.uk

1. Geijsen, N. *et al.* *Nature* **427**, 148–154 (2004).
2. Toyooka, Y., Tsunekawa, N., Akasu, R. & Noce, T. *Proc. Natl. Acad. Sci. USA* **100**, 11457–11462 (2003).
3. Hübner, K. *et al.* *Science* **300**, 1251–1256 (2003).
4. Smith, A. G. *Annu. Rev. Cell. Dev. Biol.* **17**, 435–462 (2001).
5. Matzuk, M. M., Burns, K. H., Viveiros, M. M. & Eppig, J. J. *Science* **296**, 2178–2180 (2002).
6. Saunders, P. T. *Reproduction* **61** (Suppl.), 91–101 (2003).
7. Hunt, P. A. & Hassold, T. J. *Science* **296**, 2181–2183 (2002).
8. Surani, M. A. *Nature* **414**, 122–128 (2001).
9. Saitou, M., Barton, S. C. & Surani, M. A. *Nature* **418**, 293–300 (2002).
10. *Nature* **422**, 787 (2003).



100 YEARS AGO

Some experiments have recently been made to test whether the radio-activity of radium is influenced by the continuous bombardment to which it is subjected by its own radiations. In an article in this *Journal* on radium (April 30, 1903) Prof. J. J. Thomson suggested that the radio-activity of radium may possibly depend upon its degree of concentration, and that a given quantity of radium, diffused through a mass of pitchblende, may be less than when concentrated in a small mass. In order to test this point, measurements of the radio-activity of radium bromide were made when in the solid state and when diffused throughout the mass of a solution more than a thousand times the volume occupied by the radium compound... This experiment shows that, over the range investigated, the radio-activity of radium is not influenced by its own intense radiations. E. Rutherford
From *Nature* 7 January 1904.

50 YEARS AGO

During September 21–25, 1953, a conference was held by invitation of Prof. Linus Pauling at the California Institute of Technology in Pasadena, to discuss progress in the X-ray studies of the structure of proteins (and to a lesser extent of nucleic acids). The last conference of this kind was that arranged by the Royal Society and held in London during May 1952; it had been the first to include a full-scale discussion of the new polypeptide chain configurations proposed by Pauling and Corey, especially the α -helix... The most general and fundamental concept underlying the discussions was that helical arrangements are at the basis of many important biological structures, either at the atomic level as a type of configuration for long-chain molecules, or at the molecular level as a way in which larger units of structure may naturally aggregate. Although a helical model had been proposed for the polypeptide chain by H. S. Taylor as early as 1941, ... it cannot be said that the helix as a structural principle had entered into the fundamentals of our thinking up to the time of the Royal Society conference: indeed on that occasion there was strong disagreement as to the existence of helical chains. The Pasadena conference revealed that the helix has now come into its own with a vengeance; finding helices is a game played by nearly everyone in the field.

J. C. Kendrew

From *Nature* 9 January 1954.

Ecology

Clouded futures

J. Alan Pounds and Robert Puschendorf

Global warming is altering the distribution and abundance of plant and animal species. Application of a basic law of ecology predicts that many will vanish if temperatures continue to rise.

Evidence that climate change is affecting life on Earth continues to mount^{1,2}. But how great is the threat to biodiversity? On page 145 of this issue, Thomas *et al.*³ show that global warming, projected to the year 2050, could sharply increase extinction probabilities for a sample of 1,103 species representing terrestrial regions from Mexico to Australia. If temperatures follow middle-of-the-road projections, the study suggests, about one-quarter of these species may dis-

appear — a loss that would exceed that expected from habitat destruction.

Thomas *et al.* assume that each species can persist only under a particular set of climatic conditions. This 'climate envelope', assessed by modelling current geographical distribution in relation to climatic gradients, serves to predict future distribution. As warming alters these gradients, many species are shifting towards the poles or to higher elevations, their ranges often contracting as the area of

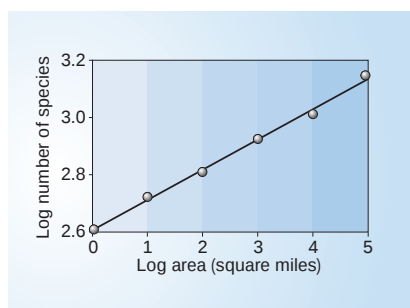


Figure 1 H. C. Watson's demonstration of the species–area law for Britain's vascular flora. The smallest sampling area was a plot in Surrey; each successively larger area was an expansion of the preceding one. Thomas *et al.*³ apply this law to assess how extinction risk increases as global warming reduces the area of climatically suitable habitat. (Adapted from ref. 4.)

climatically suitable habitat declines^{1,2}.

To predict the outcome, Thomas *et al.* turn to one of ecology's few ironclad laws: the species–area relationship. Basically, smaller areas support fewer species. In 1859, the year in which Darwin's *Origin of Species* appeared, H. C. Watson demonstrated this law for Britain's vascular flora by comparing sampling areas ranging from a square mile to all of England⁴. Plotting the logarithm of the number of species as a function of the logarithm of area, he found a linear relationship (Fig. 1). This pattern, we now know, is typical of regional scales, where a power-law equation usually describes the relationship between number of species and area⁴. At these scales, the species–area curve probably reflects the configuration of species' ranges and thus the history of speciation, dispersal and extinction.

Using this relationship to assess climate-related extinction risk, Thomas *et al.* explore three approaches. Their conclusions are the same with each of them. They consider change in area summed for the various species, proportional loss of area averaged across species, and change in area for each species individually. Averaging the results of these methods, applied under two dispersal scenarios, they estimate the extinction probability for different amounts of warming as 18% (0.8–1.7 °C), 24% (1.8–2.0 °C) and 35% (above 2.0 °C).

These estimates might be optimistic. The risk of extinction increases as global warming interacts with other factors — such as landscape modification, species invasions and build-up of carbon dioxide — to disrupt communities and ecological interactions. Furthermore, the models might not capture some key climatic changes. One pillar of Thomas and colleagues' analysis is the modelling of range contractions in the tropical rainforests of northeast Australia⁵. But the authors of that study⁵ emphasize that it considers only the effects of rising temperatures and that other changes could magnify the impacts. In the highlands, for example, an increase in the

altitudes at which clouds form^{6–8} could affect communities that require frequent immersion in clouds and mist.

Changes in cloud cover might also be important. Warming accelerates evaporation and increases the air's capacity to hold water, thereby increasing its content of water vapour⁹. Cloud formation, however, depends on relative humidity, which varies inversely with temperature, so warming may reduce cloudiness over some regions. In contrast, where air masses cool sufficiently — for instance where they ascend mountain slopes — increased water vapour should translate into enhanced cloud formation, even if condensation begins at increased altitudes. Accordingly, widespread increases in cloud cover are under way⁹.

Are changing cloud patterns already contributing to the extinction of species? Thomas *et al.*³ refer to amphibian declines and disappearances in the mountains of Costa Rica as the one example in which recent warming has been implicated in such losses (Fig. 2). Various biological changes in these mountains are associated with unusually dry weather attributed to an increase in heights of cloud formation⁶. Understanding amphibian extinctions is crucial, given that they are taking place in highlands around the world^{10,11}. For example, most of the 70-odd members of the harlequin frog genus *Atelopus*, endemic to Central and South America, have vanished or declined markedly (E. La Marca, personal communication).

Nevertheless, few studies have examined how climatic changes might be linked to the immediate causes of these declines^{12,13}. The climate-envelope concept championed by Thomas *et al.* might help to shed light on one such cause — outbreaks of the chytrid fungus *Batrachochytrium dendrobatidis*^{14,15}. This lethal parasite of amphibian skin thrives under cool, moist conditions. In culture, it grows at 6–28 °C but dies at higher temperatures. Experiments with the Australian frog *Litoria chloris*¹⁶ show that elevated body temperatures, reached naturally by basking in the sun or seeking warm microenvironments, can rid the frogs of this fungus. The low humidity typical of warm microsites might likewise enhance frog survival.

Both increased cloud cover and unusually dry weather might hamper these defences. In highland tropical forests, ambient air temperatures generally lie within the climate envelope of *Batrachochytrium*. But these forests include shaded and sunlit microhabitats. Under clear skies, temperatures in the latter can quickly exceed 30 °C, so an amphibian can 'escape' from this envelope. Under cloudy skies, however, microhabitat temperatures mirror ambient temperatures, making escape difficult. Dry conditions may have similar consequences: with limiting moisture, an amphibian might have to stay in cool, damp places.

Although exploring these potential links



Figure 2 Absent amphibians. Both the golden toad (*Bufo perigrines*, top) and the Monteverde harlequin frog (*Atelopus* sp.) were found in the mountainous Monteverde region of Costa Rica, but have not been seen since the late 1980s. New DNA evidence indicates that the latter was a yet-unnamed species (R. Ibañez, personal communication). The disappearance of these amphibians, linked to declines in mist frequency ascribed to global warming⁶, might be due to climate-related outbreaks of an emerging pathogen.

between climate and recent extinctions is essential, the patterns implicating global warming in such losses attest to the urgency of Thomas and colleagues' principal recommendation³. Reducing the concentrations of greenhouse gases — and reducing them soon — could minimize this warming and hence the number of extinctions. The threat to life on Earth is not just a problem for the future. It is part of the here and now.

J. Alan Pounds is at the Golden Toad Laboratory for Conservation, Monteverde Cloud Forest Preserve and Tropical Science Center, Santa Elena, Puntarenas 5655-73, Costa Rica.

e-mail: goldtoad@racsa.co.cr

Robert Puschendorf is in the School of Biology, University of Costa Rica, San Pedro de Montes de Oca, Costa Rica.

e-mail: rpuschen@biologia.ucr.ac.cr

1. Parmesan, C. & Yohe, G. *Nature* **421**, 37–42 (2003).

2. Root, T. L. *et al.* *Nature* **421**, 57–60 (2003).

3. Thomas, C. D. *et al.* *Nature* **427**, 145–148 (2004).

4. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).

5. Williams, S. E., Bolitho, E. E. & Fox, S. *Proc. R. Soc. Lond. B* **270**, 1887–1892 (2003).
6. Pounds, J. A., Fogden, M. P. L. & Campbell, J. H. *Nature* **398**, 611–615 (1999).
7. Still, C. J., Foster, P. N. & Schneider, S. H. *Nature* **398**, 608–610 (1999).
8. Richardson, A. D., Denny, E. G., Siccamo, T. G. & Lee, X. *J. Clim.* **16**, 2093–2098 (2003).
9. Houghton, J. T. *et al. Climate Change 2001: The Scientific Basis* (Cambridge Univ. Press, 2001).
10. Alford, R. A. & Richards, S. J. *Annu. Rev. Ecol. Syst.* **30**, 133–165 (1999).
11. Ron, S. R., Duellman, W. E., Coloma, L. A. & Bustamante, M. R. *J. Herpetol.* **37**, 116–126 (2003).
12. Kiesecker, J. M., Blaustein, A. R. & Belden, L. K. *Nature* **410**, 81–84 (2001).
13. Pounds, J. A. *Nature* **410**, 639–640 (2001).
14. Berger, L. *et al. Proc. Natl Acad. Sci. USA* **95**, 9031–9036 (1998).
15. Burrowes, P. A., Joglar, R. L. & Greene, D. E. *Herpetologica* (in the press).
16. Woodhams, D. C., Alford, R. A. & Marantelli, G. *Dis. Aquat. Org.* **55**, 65–67 (2003).

Astronomy

To catch a stellar thief

Thomas Matheson

A supernova seen in 1993 defied explanation. Astronomers suspected the dying star had lost much of its hydrogen before the explosion. The discovery of a companion to this star suggests where that gas went...

In March 1993, something new appeared in the nearby galaxy M81. A star had exploded, becoming the brightest supernova visible in the Northern Hemisphere since 1954. The exceptional brightness of SN1993J, as it is known (Fig. 1), meant that detailed observations of it were possible across virtually the entire spectrum of electromagnetic radiation. But these studies produced some unexpected results, forcing astronomers to change the way they classify supernovae. Peculiarities in the sequence of emission from SN1993J led many to conclude that the dying star must have lost a considerable amount of its outer envelope of hydrogen gas before it exploded^{1–7}; probably this mass was transferred to a companion star. Ten years later, in a dramatic vindication, Maund *et al.*⁸ report the discovery of this companion (page 129 of this issue).

That SN1993J was missing some of its hydrogen is a key factor in understanding the mechanisms that produce supernovae. Supernovae are classified according to certain features in the spectrum of radiation they produce at optical wavelengths⁹. Originally, the distinction was based solely on the presence or absence of emission from hydrogen atoms (hydrogen lines). Supernovae without hydrogen lines were called type I, those with hydrogen lines type II. Subsequent analysis of many of these objects revealed that this empirical classification scheme in fact reflected two different mechanisms for the supernova explosion. The consensus is that type I supernovae (lacking hydrogen) are the result of the thermonuclear disruption of a white-dwarf star that has absorbed almost as much mass as it can (it is near the Chandrasekhar mass limit for such stars, around 1.4 times the mass of the Sun). Type II supernovae (with hydrogen) are thought to be the result of core collapse in a massive star.

As stars develop, they begin by fusing hydrogen into helium. Once a star has exhausted its hydrogen reserves, it will then

fuse helium into carbon. Massive stars (more than eight times the mass of the Sun) can fuse carbon and heavier elements. So several stages of nuclear burning occur, giving the star a layered, onion-like structure, with an outermost shell of hydrogen surrounding shells of helium and of heavier elements. The creation of iron marks the limit for normal stellar fusion; fusing iron into heavier elements no longer releases energy. At this point, there is nothing left to counteract gravity, and the stellar core collapses to form a neutron star or, possibly, a black hole.

With bigger telescopes, better instruments and more data on supernovae, it became clear, however, that the empirical classification scheme could not remain so simple. Some supernovae that lacked hydrogen (hence type I)

did not match the homogeneous spectra expected for the thermonuclear disruption of a white dwarf. In the time-honoured astronomical tradition of inscrutable nomenclature, the original class was now renamed type Ia; the new subtypes were called type Ib (if they showed helium lines) or type Ic (if they did not). The question then arose: what mechanism created these different types of supernova?

A circumstantial case grew that type Ib and Ic supernovae were in fact related to type II supernovae: all three types are associated with massive stars and occur in similar galactic environments; a while after the explosion, their spectra look alike, except that type II supernovae show hydrogen lines. But the data for SN1993J brought another surprise¹⁰. Shortly after the explosion, its spectrum contained hydrogen lines — a type II supernova — but within a few weeks strong helium lines developed. The spectrum then looked more like a type Ib supernova, as though the envelope of hydrogen had lifted to reveal a helium layer (it was then called a IIb). A natural explanation is that these types (II, Ib and Ic) all arise from the same mechanism of core collapse, and the different spectra reflect different outermost layers for the progenitor star: the progenitors of type II supernovae retain their hydrogen; those of type Ib supernovae lose their hydrogen, but keep their helium layer; and those of type Ic supernovae have lost both hydrogen and helium layers.

If this is the case, then there must be a process by which the outer parts of the progenitors are removed before they explode. Stars can lose mass through winds (such as the solar wind of our own Sun), but this method is not efficient enough to explain the mass loss in



Figure 1 The beautiful spiral structure of the galaxy M81, one of the brightest galaxies in the sky. In 1993, a dying star in one of the arms of this galaxy exploded in a supernova, now called SN1993J. Radio-wavelength images¹³ (lower images) trace the development of SN1993J over the first year of its existence.

R. GENDLER

this instance. (Very massive stars, though, can produce very strong stellar winds; such stars may be the progenitors of type Ic supernovae, which are also associated with γ -ray bursts¹¹.) The most likely explanation for the behaviour of SN1993J is that it was part of a binary system and the missing mass was transferred from the progenitor to a companion star.

But no companion to SN1993J had been seen. Studies of images taken by the Hubble Space Telescope (HST) in 2001 were inconclusive on the presence of a star among the material ejected by the supernova¹². In May 2002, however, Maund *et al.*⁸ were able to take advantage of the exquisite resolution available with the new Advanced Camera for Surveys on the HST. They report an excess of light coincident with the ejecta, implying that there is another star present. In addition to these photometric measurements, Maund *et al.* gathered spectroscopic data with the Keck I telescope in Hawaii: the brightness of SN1993J has dimmed sufficiently that its spectrum now shows the features of a star superimposed on the supernova. It seems that the thief that stole most of the hydrogen from the progenitor of SN1993J has finally been captured.

The spectrum from the ground-based Keck I telescope does, however, include some light from neighbouring stars, owing to the blurring effect of the Earth's atmosphere.

When SN1993J has faded more, the detection of the companion star should become absolutely certain. It will then take careful study to establish whether the companion really did acquire the hydrogen from SN1993J's progenitor. Nevertheless, observations strongly suggest that type Ib and Ic supernovae are linked to core collapse inside stars. Further data should confirm whether their distinctive spectra are indeed caused by the theft of the outer layers of a dying star by a merciless companion. ■

Thomas Matheson is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.
e-mail: tmatheson@cfa.harvard.edu

1. Nomoto, K. *et al.* *Nature* **364**, 507–509 (1993).
2. Podsiadlowski, Ph. *et al.* *Nature* **364**, 509–511 (1993).
3. Ray, A., Singh, K. P. & Sutaria, F. K. *J. Astrophys. Astron.* **14**, 53–63 (1993).
4. Shigeyama, T. *et al.* *Astrophys. J.* **420**, 341–347 (1994).
5. Woosley, S. E. *et al.* *Astrophys. J.* **429**, 300–318 (1994).
6. Bartunov, O. S. *et al.* *Astron. Astrophys.* **281**, L53–L55 (1994).
7. Utrobin, V. *Astron. Astrophys.* **281**, L89–L92 (1994).
8. Maund, J. R., Smartt, S. J., Kudritzki, R. P., Podsiadlowski, Ph. & Gilmore, G. F. *Nature* **427**, 129–131 (2004).
9. Filippenko, A. V. *Annu. Rev. Astron. Astrophys.* **35**, 309–355 (1997).
10. Filippenko, A. V. & Matheson, T. Preprint at <<http://arxiv.org/astro-ph/0310228>> (2003).
11. Stanek, K. Z. *et al.* *Astrophys. J.* **591**, L17–L20 (2003).
12. Van Dyk, S. D. *et al.* *Pub. Astron. Soc. Pacif.* **114**, 1322–1332 (2002).
13. Marcaide, J. M. *et al.* *Science* **270**, 1475–1478 (1995).

Cancer

Guarding the guardian?

Henning F. Horn and Karen H. Vousden

The diversity of cellular stresses that activate the 'guardian of the genome' — p53 — begs the question of how these signals all converge on one protein. Perhaps the key to this integration is the nucleolus.

If there is beauty in simplicity then the p53 protein is unlikely to win any contests. As one of our key defences against cancer, p53 lies at the heart of a massive research effort that has generated an avalanche of papers. The result is a bewildering complexity to virtually every aspect of p53 biology that leaves outsiders — and many within the field — bemused and confused. Now, a paper from Rubbi and Milner¹ in *EMBO Journal* offers some hope of rationalizing the pathways that regulate p53.

The p53 protein is activated in response to cellular stresses — such as low oxygen levels, heat shock, nucleotide depletion and DNA damage — and acts to prevent further proliferation of the stressed cell by promoting cell-cycle arrest or cell suicide. These events, in turn, help to prevent cancer. An important step in the activation of these functions is the stabilization of the protein, which results in a rapid rise in cellular p53 levels. There is good evidence that the stabilization of p53 is associated with inhibition of HDM2 in humans

(Mdm2 in mice), a protein that, if not kept in check, promotes the destruction of p53. But how the cell integrates and interprets so many stress signals from such diverse stimuli to activate p53 has been an enigma. Rubbi and Milner¹ propose that, in fact, all these stress signals feed into p53 in the same way — by disrupting nucleoli.

Nucleoli are subnuclear structures that are important for the assembly of ribosomal subunits (ribosomes being the cellular protein-making machines). Rubbi and Milner draw on the large body of available evidence to point out that many of the known p53-activating signals can also promote nucleolar disruption. The authors also find that when they focus ultraviolet radiation on small sections of the nucleus, in such a way as to cause DNA damage but not compromise nucleolar function, p53 does not become stabilized. Conversely, injecting an antibody that interferes with nucleolar function — causing nucleolar disruption without DNA damage — does lead to p53 stabilization. The inter-

esting conclusion is that DNA damage *per se* is not enough to stabilize p53, but that stabilization depends on nucleolar disruption.

It is clear that many of the abnormalities in cell growth and proliferation that are associated with cancer development are likely to cause nucleolar stress, and it makes sense that this should trigger a protective response by p53. But Rubbi and Milner's proposal — that all p53-inducing signals are actually transmitted through the nucleolus — is rather bold and exciting. Early studies strongly suggested that DNA damage itself signals to p53, with the calculation that a single unrepaired DNA strand break is sufficient to activate a p53 response². Moreover, a compelling case for a direct role of DNA strand breaks was made by the observation that injecting the restriction enzyme *PvuII*, which cuts DNA, could stabilize the p53 protein^{3,4}. Although these treatments might not be anticipated to induce extensive nucleolar disruption, it will be important to revisit these findings in the light of the new information.

Another issue raised by Rubbi and Milner is just how nucleolar disruption could lead to p53 activation. There are efficient cellular mechanisms that export ribosomal subunits from the nucleus to the cytoplasm — where protein synthesis occurs — and it has been proposed that p53 uses these same mechanisms for its own export⁵. Moreover, some (if not all) studies have linked p53 export to its degradation. If so, there is some logic to Rubbi and Milner's suggestion that disruption of nucleoli may lead to p53 stabilization by perturbing the p53-export mechanism.

In this regard it is interesting that nucleophosmin — a nucleolar protein that has been proposed to shuttle proteins from the nucleolus to the cytoplasm — can also help modulate p53 stability in response to various forms of stress⁶. Interestingly, nucleophosmin appears to play a role in allowing p53 stabilization, although whether this reflects a disruption of the normal shuttling of p53 has not yet been investigated. Several other nucleolar proteins have also been shown^{7–9} to bind p53, but again, the consequences of these interactions for p53 export are not fully established.

Indeed, it is not even certain whether the regulation of p53 shuttling is the key to p53 activation. An alternative possibility is that nucleolar disruption stabilizes p53 by releasing specific nucleolar proteins into the body of the nucleus. These might include ribosomal proteins, such as L11, that influence p53 stability by binding to and inactivating HDM2 (refs 10, 11). Or they could include some of the other nucleolar proteins that bind p53 directly. Interestingly, a role for the stress-induced release of nucleolar proteins in carrying out other regulatory functions is becoming something of a general theme. Recent examples include the interferon- γ -induced release of the ribosomal protein L13a (ref. 12), which leads to the silencing of

protein synthesis, and the inhibition of DNA replication following stress-induced release of the protein nucleolin⁸.

There has been a remarkable convergence of recent evidence — including the Rubbi and Milner paper¹ — suggesting that nucleoli are important in monitoring cellular stress. The health of the nucleolus is an excellent surrogate for the health of the cell, and conditions that lead to nucleolar disruption are unlikely to be safe for continued cell proliferation. The notion that intact nucleoli are necessary to hold the p53 response in check provides an attractive model in which a default pathway of p53 induction and inhibition of cell growth is overcome only by the maintenance of nucleolar well-being. These ideas reinforce the growing realization that the nucleolus — long regarded as a mere factory for assembling ribosomal subunits — is a vital command unit in

monitoring and responding to stress.

Henning F. Horn and Karen H. Vousden are at the Beatson Institute for Cancer Research, Switchback Road, Glasgow G61 1BD, UK.

e-mail: k.vousden@beatson.gla.ac.uk

1. Rubbi, C. P. & Milner, J. *EMBO J.* **22**, 6068–6077 (2003).
2. Leonardo, A. D., Linke, S. P., Clarkin, K. & Wahl, G. M. *Genes Dev.* **8**, 2540–2551 (1994).
3. Siegel, J., Fritsche, M., Mai, S., Brandner, G. & Hess, R. D. *Oncogene* **11**, 1363–1370 (1995).
4. Lu, X. & Lane, D. P. *Cell* **75**, 765–778 (1993).
5. Sherr, C. J. & Weber, J. D. *Curr. Opin. Genet. Dev.* **10**, 94–99 (2000).
6. Colombo, E., Marine, J.-C., Danovi, D., Falini, B. & Pelicci, P. G. *Nature Cell Biol.* **4**, 529–533 (2002).
7. Tsai, R. Y. & McKay, R. D. *Genes Dev.* **16**, 2991–3003 (2002).
8. Daniely, Y., Dimitrova, D. D. & Borowiec, A. *Mol. Cell. Biol.* **22**, 6014–6022 (2002).
9. Blander, G. *et al.* *J. Biol. Chem.* **274**, 29463–29469 (1999).
10. Lohrum, M. A. E., Ludwig, R. L., Kubbutat, M. H. G., Hanlon, M. & Vousden, K. H. *Cancer Cell* **3**, 577–587 (2003).
11. Zhang, Y. *et al.* *Mol. Cell. Biol.* **23**, 8902–8912 (2003).
12. Mazumder, B. *et al.* *Cell* **115**, 187–198 (2003).

studies contains significant gaps. The paper by Raya *et al.*² on page 121 of this issue goes some way towards completing this picture, revealing an explicit link between an early, temporary asymmetry and later, stable patterns of asymmetric gene expression.

The events that lead to the initial breaking of left–right symmetry in vertebrate embryos are not fully understood, but they are believed to provide only weak transient biases³. So additional mechanisms must exist to amplify these biases, converting them into stable and heritable asymmetric patterns of gene expression¹. The earliest detected feature of left–right asymmetry that is common to all vertebrates studied is the expression of the secreted growth-factor protein Nodal on the left side of the ‘node’. This region, located on the midline of the embryo, acts as an organizing centre during development. In mice, Nodal expression has been shown to depend on a second signalling pathway, centred on the cell-surface-located receptor Notch^{4,5}. But how the Notch pathway becomes activated to a sufficient degree to trigger Nodal expression only on the left side of the node remains an open question.

Raya *et al.*² use a combination of modelling and experimentation to address this problem in chick embryos. Having determined the patterns of expression of various key genes around the node, the authors capitalize on this information to construct a mathematical model of the network of molecular interactions underlying Notch activation and Nodal expression. As Nodal enhances its own production, it can act as an on–off switch: only a transient increase in activity of the Notch pathway is required to induce stable Nodal expression. Raya *et al.* find that the simplest way to achieve this in their model is to enhance the affinity of Notch for its activating partners (ligands) — the Delta-like 1 (Dll1) and Serrate 1 (Srr1) proteins. So the model suggests that a transient lateral bias in this affinity should be enough to convert the initially symmetric pattern of gene expression into one that is manifestly asymmetric.

The authors carry out a range of experiments that show that this is indeed the case. In doing so, they uncover a chain of events that lead from a left–right asymmetry in the electrochemical potential across the membranes of cells around the node, to the left-specific expression of Nodal. The first step in this cascade is a previously described left-sided reduction in the activity of a membrane-spanning ion pump (the H⁺/K⁺-ATPase); this reduction results in membrane depolarization⁶. Raya *et al.* find that this depolarization leads to a transient increase in the extracellular concentration of Ca²⁺ ions on the left of the node. And this in turn is necessary for left-sided Nodal expression — suggesting that it could be Ca²⁺ that modulates the affinity of Notch for its ligands. In

Developmental biology

Asymmetric fixation

Nick Monk

Computer simulations and laboratory experiments have shed light on how an asymmetric pattern of gene expression is fixed in vertebrate embryos — an early step towards asymmetric development of the internal organs.

As judged by external appearances, the left and right sides of vertebrate bodies are (more or less) identical. There are, however, consistent left–right differences in the structure and placement of the internal organs. The heart, for instance,

usually forms on the left, the liver on the right. In recent years, researchers have uncovered several different molecular events that are involved in establishing this left–right asymmetry as embryos develop¹. But the picture that has emerged from these

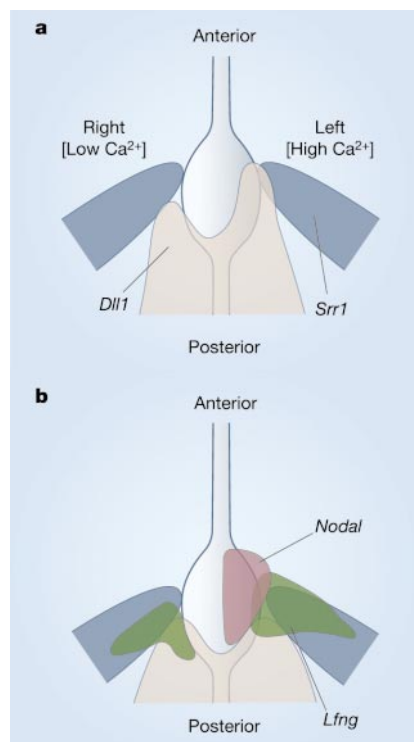


Figure 1 Fixing asymmetry in vertebrates.

According to convention, embryos are viewed from the ‘front’ — so the left-hand side of the embryo appears on the right of this diagram. An early manifestation of asymmetry in chick embryos is the expression of the *Nodal* gene on the left of the ‘node’ (oval). Raya *et al.*² put forward a model for how this occurs. It was known from studies in mice that *Nodal* expression depends on the Notch pathway, which is in turn activated by Dll1 and Srr1. a, At stage 5 of development (19–22 hours after fertilization), *Dll1* expression extends further towards the head (the anterior) on the left than on the right. This is the earliest indication that Notch activity is higher on the left (as *Dll1* is a target of Notch activity). b, During stage 6 (23–25 hours after fertilization), the *Dll1* and *Srr1* expression domains are symmetrical. But, as the fifth pulse of expression of the *Lfng* gene sweeps up the embryo, it moves further to the anterior on the left. *Nodal* expression is then induced around the boundary between *Dll1* and *Srr1* expression. This occurs only on the left, where the Ca²⁺ concentration is high; this might enhance the affinity of Notch for its ligands. Note that the node ‘regresses’ posteriorly between stages 5 and 6.

Plant development

The flowers that bloom in the spring

Deciding when to flower is of crucial importance to plants; every season has advantages and disadvantages, and different plant species adopt different strategies. Elsewhere in this issue, Sibum Sung and Richard M. Amasino (*Nature* **427**, 159–164; 2004) and Caroline Dean and colleagues (*Nature* **427**, 164–167; 2004) investigate how such decisions are made at the molecular level. They uncover a mechanism that prevents the model plant *Arabidopsis thaliana* (pictured) from blooming until the coming of spring.

Plants take a variety of environmental factors into account when choosing when to flower, such as the length of the day, the plant's age and the requirement for an extended cold period (a process called vernalization). All of these factors work in part through the gene *FLOWERING LOCUS C* (*FLC*),

whose protein product blocks flowering by repressing numerous genes required for flower development. During a prolonged cold spell, for example, the normally high levels of expression of *FLC* are lowered, remaining low even after warm weather returns.

Several genes are needed for vernalization: Dean and colleagues studied two of these, *VRN1* and *VRN2*, whereas Sung and Amasino identified another, *VIN3*. All three encode proteins with counterparts in animals that either bind DNA directly, or change the structure of the chromatin into which DNA is packaged.

Following this lead, the two groups found that vernalization induces changes in histone proteins (components of chromatin) in the vicinity of the *FLC* gene — and that *VRN1*, *VRN2* and *VIN3* mediate these



changes. Specifically, cold causes the loss of acetyl groups from particular lysine amino acids in histone H3. Such patterns of deacetylation mark genes that are permanently inactivated or silenced. The researchers found that whereas *VIN3* is needed to deacetylate H3 during a cold snap, *VRN1* and *VRN2* are required afterwards, to maintain the silenced state.

Interestingly, these changes in histone acetylation are confined to a region of the *FLC* gene that was recently shown to contain a binding site for the FLOWERING LOCUS D (FLD) protein (Y. He *et al. Science* **302**, 1751–1754; 2003). FLD is related to a component of the human histone deacetylase complex, and is also involved in promoting flowering by silencing *FLC*. Plants lacking FLD show both high levels of histone acetylation and a considerable reluctance to flower.

Silencing is an effective means of controlling long-term gene expression, as it persists even after cells divide. In animals, switching silencing on or off is a well-known way to control development. It seems that plants share this system, using it to preserve the memory of winter's passing. **Christopher Surridge**

support of this, the authors discover that ligand-dependent activation of Notch in cultured cells is sensitive to Ca^{2+} concentrations in the range observed around the chick node.

These findings provide a convincing picture of how Notch can trip the Nodal switch asymmetrically. The Nodal gene is, however, expressed only in a restricted region immediately neighbouring the node (Fig. 1), whereas the Ca^{2+} concentration increases in a much broader domain. Raya *et al.* show that this spatial restriction depends on a second input to the Notch pathway. The Notch ligands Dll1 and Srr1 are expressed on both the left and right of the node, in regions that abut at an interface that lies roughly perpendicular to the embryo's head-to-tail axis. It is around this interface on the left of the node — where Ca^{2+} levels are high — that Nodal is expressed (Fig. 1). This is not a coincidence: Raya *et al.* find that experimentally disrupting this interface results in loss of left-sided Nodal expression.

A third input is required to determine the time at which the Notch pathway turns on Nodal expression. Raya *et al.* show that the Lunatic fringe (Lfng) protein is an essential component of this input. The expression of this protein is highly dynamic — several short pulses of Lfng expression sweep up the embryo from tail to head⁷. Raya and colleagues' findings suggest that, as these pulses cross the Dll1–Srr1 interface, they enhance Notch activation. On the left of the node, where Notch activity is already higher than on the right because of the asymmetry in

Ca^{2+} levels, the fifth wave of Lfng expression raises Notch activity to a high enough level to allow Nodal to be expressed (Fig. 1).

This work represents a significant advance in our understanding of how left–right asymmetry is established. It shows for the first time how transient non-genetic biases can become fixed in stable asymmetric patterns of gene expression. It also provides a concrete example of a patterning mechanism that is driven by the spatial modulation of a kinetic parameter (the affinity of Notch for its ligands)⁸. A central role is played by the Notch pathway, which acts as a robust signal integrator and amplifier, using three disparate inputs to ensure that Nodal is expressed at the correct time and place. Raya and colleagues' approach illustrates the benefits that can be gained by exploiting the complementarity of theoretical and experimental approaches, especially in systems as complex as vertebrate embryos.

There are, of course, a few gaps yet to fill. Most obviously, how is left–right symmetry broken in the first place? In mice, an attractive candidate for the symmetry-breaking event is the right-to-left flow of extracellular fluid seen around the node⁹. The motile cilia that generate this flow have been observed in several different vertebrates before left-sided Nodal expression is established, prompting speculation that fluid flow has an evolutionarily conserved role in generating left–right asymmetry^{10,11}. But expression of Notch around the node and fluid flow (or its consequences) appear to be largely independent of

each other^{4,5}. It is intriguing that fluid flow also generates a brief increase in Ca^{2+} levels to the left of the node — although this rise is intracellular rather than extracellular¹². Perhaps these seemingly parallel mechanisms are somehow integrated at the level of Nodal expression.

There are further issues. How does the juxtaposition of Dll1 and Srr1 expression enhance Notch activity? How is this potentiated by Lfng? And are there parallels with the activation of Notch at Fringe-demarcated boundaries in fruitflies? The dramatic progress made in recent studies has opened up many new fronts on which to explore these fascinating questions. ■

Nick Monk is at the Centre for Bioinformatics and Computational Biology and in the Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield S1 4DP, UK. e-mail: n.monk@shef.ac.uk

- Hamada, H., Meno, C., Watanabe, D. & Saijoh, Y. *Nature Rev. Genet.* **3**, 103–113 (2002).
- Raya, A. *et al. Nature* **427**, 121–128 (2004).
- Mercola, M. J. *Cell Sci.* **116**, 3251–3257 (2003).
- Krebs, L. T. *et al. Genes Dev.* **17**, 1207–1212 (2003).
- Raya, A. *et al. Genes Dev.* **17**, 1213–1218 (2003).
- Levin, M., Thorlin, T., Robinson, K. R., Nogi, T. & Mercola, M. *Cell* **111**, 77–89 (2002).
- Jouve, C., Iimura, T. & Pourqu  , O. *Development* **129**, 1107–1117 (2002).
- Page, K. M., Maini, P. K. & Monk, N. A. M. *Physica D* **181**, 80–101 (2003).
- Nonaka, S. *et al. Cell* **95**, 829–837 (1998).
- Essner, J. J. *et al. Nature* **418**, 37–38 (2002).
- McGrath, J. & Brueckner, M. *Curr. Opin. Genet. Dev.* **13**, 385–392 (2003).
- McGrath, J., Somlo, S., Makova, S., Tian, X. & Brueckner, M. *Cell* **114**, 61–73 (2003).

Neurobiology

Conservative tastes in *Drosophila*

Development 131, 83–92 (2004)

In insects that undergo metamorphosis, the transition from a creeping larva to a flying adult is accompanied by a radical remodelling of the nervous system. It is generally accepted that all the sensory organs die during this process and are rebuilt from scratch, using cells derived from structures known as imaginal discs. However, new findings in the fruitfly *Drosophila melanogaster* challenge this view.

Nanaë Gendre and colleagues followed the fate of individual sensory neurons from their time of birth in the larva through to adulthood in the fly. They found that one gustatory (taste) organ in the larval pharynx survives metamorphosis virtually intact, and another splits to generate two gustatory organs in the adult pharynx.

Why do these particular sense organs survive? Gendre *et al.* suggest that they might be required during metamorphosis, or that it might simply be more economical to retain structures whose functions are conserved between the larva and the adult. The underlying developmental processes are also intriguing. For example, it will be interesting to discover what factors contribute to the selective survival of these sensory neurons and how they become wired into the adult nervous system.

Heather Wood

Microelectronics

A wet transistor

Appl. Phys. Lett. 83, 5310–5312 (2003)

Transistors that use liquid electrolytes as an active element have been shrunk to the nanoscale by Z. G. Chiragwandi and colleagues. Larger-scale versions of these electrolyte transistors have been proposed previously as electrochemical sensors and as biocompatible devices for use in physiological environments.

Chiragwandi *et al.* have used electron-beam lithography to make a silicon-based device, with electrodes just 600 nm square, that amplifies a current when a droplet of pure, deionized water is placed on top. It is a bipolar transistor: a three-terminal device in which, conventionally, the current passing between an 'emitter' and a 'collector' electrode is controlled by the voltage between the emitter and the third 'base' electrode.

In this water-based device, the current is conducted through the water by ions, and the degree of water ionization is altered by the base-emitter voltage. In other words, this voltage changes the hydrogen-ion

concentration (pH) even though the ratio of H^+ to OH^- ions stays equal. So the pH determines the current through the device. The authors say that the transistor should work even in salty biological media.

Philip Ball

Cancer

Sheep in wolf's clothing

Cancer Cell 4, 477–482 (2003)

By definition an oncogene promotes tumour formation. But Alexander Fleischmann and colleagues find that switching off one such gene, *Fos*, can also stimulate tumour growth under specific circumstances. In other words, *Fos* can behave as a tumour suppressor as well as an oncogene.

The authors inactivated *Fos* in mice that already lacked the *Trp53* gene. *Trp53* encodes a tumour suppressor that works by blocking cell division and inducing cell death. It is frequently mutated in human cancers, and without it mice develop a wide spectrum of malignancies.

Fleischmann *et al.* find that without both *Fos* and *Trp53*, mice rapidly form tumours in the head and neck region. The pathology closely resembles the most common variant of human rhabdomyosarcoma — a childhood cancer that arises from unbridled growth of primitive muscle cells. So the authors speculate that inactivation of *Fos* might contribute to this human cancer. They further show that reintroducing *Fos* into tumour cells *in vitro* causes many of the cells to die, suggesting that *Fos* normally suppresses tumour growth by activating the cellular suicide programme.

The mice without *Trp53* and *Fos* might provide the best animal model yet of rhabdomyosarcoma. And whether *Fos* acts as an oncogene or a tumour suppressor seems to depend strictly on its context.

Marie-Thérèse Heemels

Developmental biology

Stem cells profiled

PLoS Biol. 1, 410–419 (2003)

Stem cells can turn into many different cell types, raising hopes that they could be used to repair damaged tissues and organs. Little is known, however, about the molecular signals that guide their conversion.

Alexei A. Sharov *et al.* have now compiled a database of 30,000 genes that are involved in early embryonic and stem-cell development in mice. The team combined new and old gene-expression data from a wide range of cell types and developmental stages; in so doing they discovered 977 previously unidentified genes. One set of 88 genes might serve as a molecular marker of

developmental potential — the expression of these genes decreases from eggs to early embryos to newborns, as the cells found at these stages become progressively more restricted in the number of cell types they can generate.

The database offers a preliminary set of molecular markers with which to characterize the function and potential of different stem cells. The authors hope that the catalogue will speed research in reproductive and regenerative medicine.

Helen R. Pilcher

Planetary science

Lava at Loki

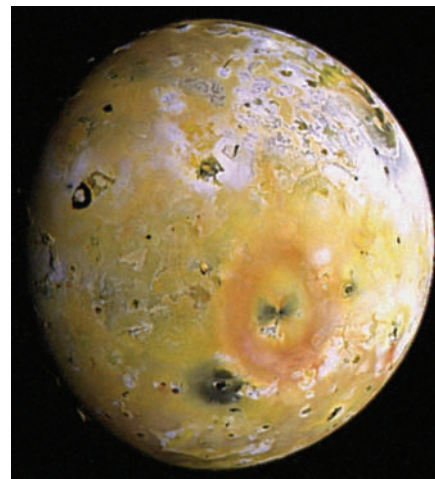
Geophys. Res. Lett. doi: 10.1029/2003GL018371 (2003)

Io, a satellite of Jupiter, seethes with volcanic activity. One of the hottest of Io's hotspots is Loki Patera (see picture), a huge area some 180 km across, which in 2001 was subject to close inspection by instrumentation on the Galileo spacecraft.

Ashley Gerard Davies has used those data to put some figures on a then-fresh bout of volcanism at Loki. A temperature profile along a lava flow can be converted into age by using a cooling curve, a calculation that also produces an estimate of the thickness of the flow. Lava was being produced from the southwest corner of the feature at a rate of about 1 km per day. Age along the flow ranged from 10 to 80 days at the furthest extent, and thickness from about 2.6 m at the source to 0.9 m.

But what is the underlying mechanism? Loki could be a massive lava lake, with the balanced creation and foundering of crust from a lava supply deep within Io. Or it could be the result of a confined, 'ponded' flow, the features of which are more a consequence of topography than the plumbing arrangements. What makes Loki tick, says Davies, remains an open question.

Tim Lincoln



An image of Io, taken by Galileo in 1997. One large volcanic centre, Pele, is at the centre of the red circle. Loki Patera is the black horseshoe-shaped feature towards the upper left.

NASA

Explosive craters and soil liquefaction

Curious dry craters formed in the aftermath of a disastrous earthquake are explained.

The Bhuj earthquake in January 2001, with a moment magnitude of 7.7, was one of the most disastrous in India's recorded history. Although its fault plane did not rupture the surface, the earthquake caused ground failure, including the formation of sand-blow craters and sand fissures, indicative of liquefaction of subsurface sediment, over an area of about 15,000 square kilometres¹. Here we investigate the two large, dry craters that formed during the earthquake, and estimate the likely subsurface pressure at the time from a trajectory analysis of the ejected clasts. We conclude that the craters were created by explosive deformation of the ground that probably resulted from the delayed effects of soil liquefaction.

Liquefaction is a result of the cyclic loading of the ground by intense shear waves during a severe earthquake, leading to an increase in pore-water pressure and a reduction in the strength of water-saturated sediments^{2,3}. At a liquefaction site near the town of Umedpur, about 48 km west-northwest of the earthquake's epicentre (Fig. 1a), two remarkable dry craters (2.4 × 1.8 m and 1.6 × 1.5 m) formed, as well as several large sand-blow craters and sand fissures (Fig. 1b). The field of clasts ejected from the dry craters suggests that they were formed by an explosive ground failure. The development of craters has been attributed to large hydraulic gradients in buried clay layers⁴. Inspection revealed that vented sand and silt were already on the surface when these craters formed, as shown by the impact/splash features created by the ejected clasts (Fig. 1c).

We estimated the subsurface pressure during the earthquake from an analysis of the calculated trajectories for clasts of different sizes. The impact distance of a thrown object with take-off angle θ and velocity v is $x_{\max} = (2v^2 \sin\theta \cos\theta)/g$, where g is the acceleration due to gravity. If the pressure during the explosion acts similarly on all clasts, the smaller, less massive clasts will attain a greater velocity and travel farther than the larger clasts (Fig. 1b, d). The velocity depends on the pressure, P , that provides the force to accelerate, a , the clasts. We assume that the pressure is constant for the duration, τ , of the ejection, which gives $v = a\tau = P\tau/\rho L$, where ρ is the density and L is the size of a clast. Substituting v in the distance formula gives a relation between clast size and impact distance of $L^2 = \kappa(P\tau)^2/(1/x_{\max})$, where the constant $\kappa = (2\sin\theta \cos\theta)/\rho^2 g$.

A plot of L against $1/x_{\max}$ will be a straight line with a slope related to $P\tau$. Although there is a considerable scatter in clast size and distance, a rough estimate is obtained by taking

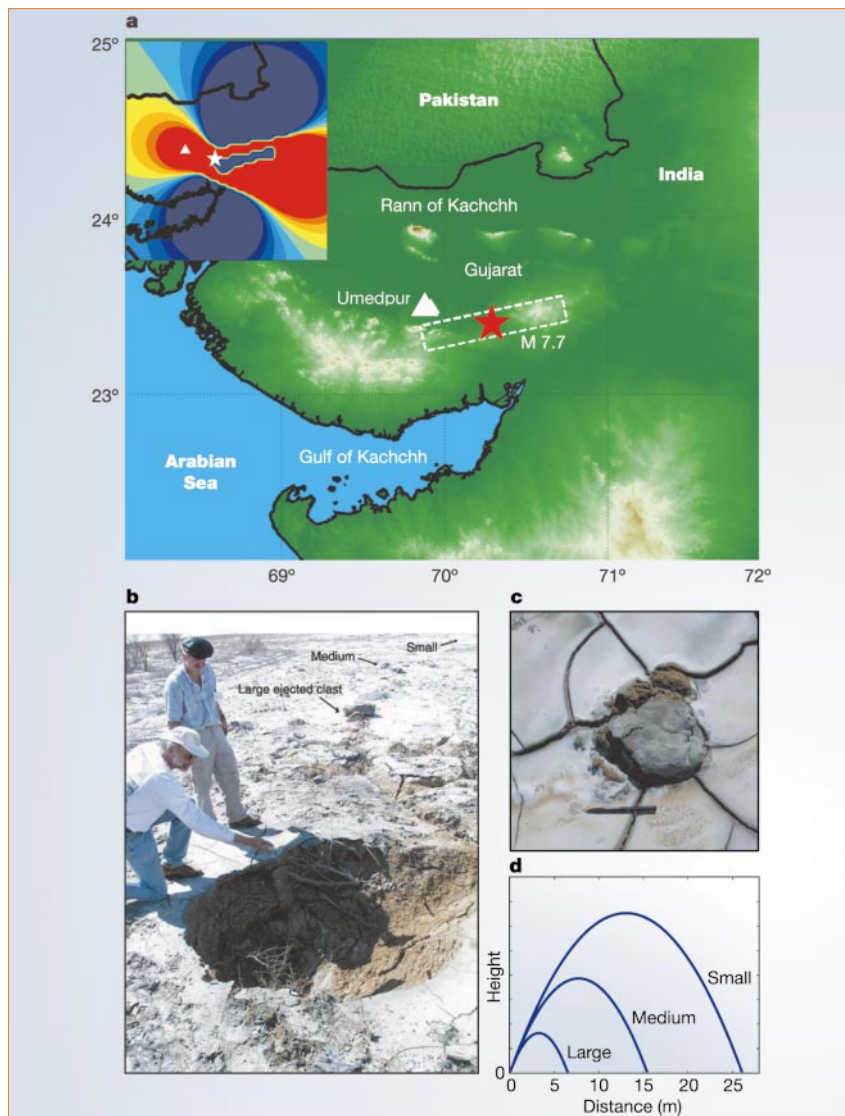


Figure 1 Explosive ground deformation from the delayed effects of soil liquefaction during the Bhuj earthquake (magnitude 7.7) on 26 January 2001. **a**, Map of northwestern India showing the location of the earthquake's epicentre (star) and the Umedpur liquefaction site (triangle); inset, contour map showing the co-seismic volumetric strain (red, compression; blue, rarefaction). **b**, Photograph of a dry crater and the ejected clasts. **c**, Close-up of an ejected clast showing the deformation of the surrounding sediment as a result of the impact (pen is shown for scale). **d**, Trajectories for ejected clasts of different sizes. Maximum trajectory height is 6.5 m. From the size and distance of the thrown clasts, the explosive pressure can be estimated by using trajectory theory.

spatial-average measurements at several impact distances up to a maximum of 26 m. From the craters' geometry, we estimated the take-off angle, θ , as being about 45°.

From the slope of the straight-line fit to our data, we obtain an estimate for $P\tau$ of 11 kPa s. Guided by villagers' accounts that a loud report was heard, we assume that the explosion occurred in about 0.2 s, giving a pressure of 56 kPa, or the equivalent of a 5.6-m column of water. This simple model ignores the dynamic interaction between gas pressure

and the break-up and ejection of the clasts, and so probably represents an underestimate of the actual explosive pressure. Our result is consistent, however, with reports of water-fountain heights of 3–4 metres seen soon after the earthquake.

Crater formation at Umedpur occurred in two phases. As such, the dry craters, which formed after the large sand-blow craters and fissures, are remarkable examples of the delayed effects of liquefaction. The exact mechanism of formation cannot be

conclusively determined, but several possible models could account for this delay. The difference in timing may be related to a variability in the subsurface stratigraphy and/or to liquefaction in more than one layer. Conditions may have been such that a relatively impermeable layer allowed the development of a pressure head that was ultimately released explosively, giving rise to the craters (deeper and thinner clay layers develop greater maximum hydraulic gradients⁴). Dry craters would be generated by the release of large bubbles of gas, perhaps as a result of aggregative fluidization⁵ or of the collapse of a gas-filled cavity at depth soon after liquefaction occurred, which would force gas to the surface.

The severity of ground failure as a result of liquefaction is related to site conditions, ground motions and distance from the epicentre⁵. In general, Bhuj's liquefaction features decrease in size with increasing epicentral distance. However, some of the

largest features occur west of the fault rupture, possibly as a result of western directivity of seismic energy^{1,6}. As-yet-unknown soil conditions and ground motions at the site are required to model the liquefaction potential, for which the subsurface pressure estimated here could prove a key parameter.

Paul A. Rydelek*, Martitia Tuttle†

*Center for Earthquake Research and Information, University of Memphis, Tennessee 38152, USA
e-mail: par@ceri.memphis.edu

†M. Tuttle and Associates, 128 Tibbetts Lane, Georgetown, Maine 04548, USA

1. Tuttle, M. P. et al. *Earthquake Spectra* **18**, 79–99 (2002).
2. Allen, J. R. L. *Sedimentary Structures: Their Character and Physical Basis* (Elsevier, Amsterdam, 1982).
3. Seed, H. B. & Idriss, I. M. *Ground Motions and Soil Liquefaction During Earthquakes* (Earthquake Eng. Res. Inst., Univ. California, Berkeley, 1982).
4. Elgamal, A.-W., Dobry, R. & Adalier, K. in *Proc. 2nd US–Japan Workshop on Liquefaction, Large Ground Deformation and their Effects on Lifelines* 233–245 (NCEER-89-0032, 1989).
5. Bartlett, S. F. & Youd, T. L. NCEER-92-0021 (Buffalo, NY, 1992).
6. Bendick, R. et al. *Seism. Res. Lett.* **72**, 328–335 (2001).

Competing financial interests: declared none.

Acoustics

Tuning of vocal tract resonance by sopranos

Sopranos can sing at frequencies that are rather higher than the normal values for the lowest resonance of their vocal tract, but failure to use this resonance would reduce both their vocal power and homogeneity in timbre. We have directly measured the resonance frequencies of the vocal tract of sopranos during singing, and find that, towards the top of their range, they consistently increase the frequency of the lowest resonance to match that of their singing. This significantly increases the loudness and the uniformity of tone, albeit at the expense of comprehensibility.

The different vowel sounds in normal speech are produced by adjusting the position of the tongue, jaw and lips so that the vocal tract resonates at certain specific frequencies (R_i). The vocal folds vibrate at the pitch frequency, f_0 , which is usually lower than the R_i . Those harmonics of f_0 that fall near the R_i produce associated peaks, or formants (F_i), in the speech spectrum¹. To oversimplify a little, in Western languages vowels occupy characteristic positions on a 'map' of (F_2, F_1) or (R_2, R_1).

Singers who are trained in the Western classical tradition need to make themselves heard in large auditoria, sometimes with loud accompaniment. In the high soprano range, f_0 enters the range at which human hearing sensitivity is greatest, so tuning R_1 to match f_0 is a possibility that could produce a sound with a loudness and timbre that vary less with pitch, and which is louder for constant effort^{2,3}. As vowel identifiability is inevitably compromised when f_0 greatly exceeds R_1 , this further loss of comprehensibility is tolerable. Indeed, sopranos are often taught to lower

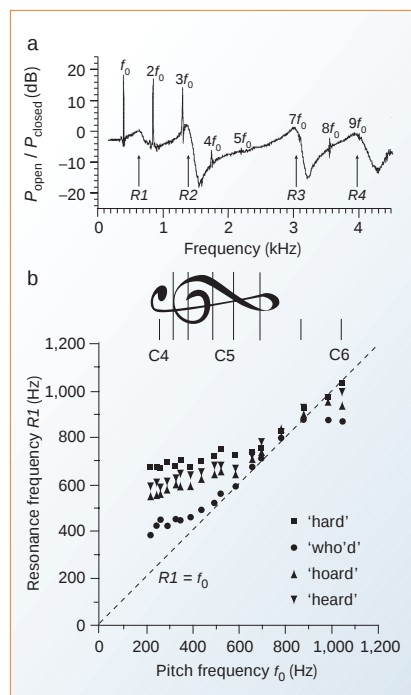


Figure 1 Simultaneous measurements of the resonance frequencies of the vocal tract and the harmonics present in the voice signal. **a**, The ratio of the spectrum measured with the mouth open, to that with the mouth closed (P_{open}/P_{closed}) when the vowel in the word 'hard' is sung at A4. Several harmonics of the voice signal with fundamental frequency $f_0 = 440$ Hz can be seen. The maxima in the broad band signal corresponding to the resonances R_1, R_2, R_3 and R_4 are indicated by arrows. **b**, The lowest resonance frequency, R_1 , as a function of the pitch frequency f_0 . The pitch is indicated in musical terms by the rotated treble clef at the top of the figure.

the jaw or to 'smile' as they ascend a scale; both of these actions increase mouth opening, which in turn increases R_1 (refs 4, 5).

We measured the resonances of the tract directly by using a previously described technique^{6,7}. Briefly, an acoustic current source

(diameter, 10 mm), synthesized from frequencies spaced at 5.38 Hz over the range 0.2–4.5 kHz, and a microphone (diameter, 8 mm) are placed just below the singer's mouth, in contact with the lower lip. The acoustic current is adjusted so that the pressure spectrum, P_{closed} , is independent of frequency when the singer's mouth is closed. The ratio of the spectrum measured with the mouth open to that with the mouth closed shows peaks that indicate the resonances of the tract⁶ as well the harmonics present in the voice signal (Fig. 1a). The shift in resonance frequency due to the reduction in area for radiation caused by the presence of the apparatus lay within the resolution of these experiments (± 11 Hz).

Eight sopranos (four professional, four advanced students), with an average of nine years' classical training, sang notes sustained for 4 s without vibrato in an ascending diatonic scale indicated by a glockenspiel before each note. They sang one of four vowels presented in writing in the form 'h<vowel>d' ('hard', 'hoard', 'who'd', 'heard'). They sang *piano* (softly) to avoid saturating the microphone and to improve the signal-to-noise ratio.

When f_0 was less than the value of R_1 for normal speech, the resonance for each vowel was roughly constant (Fig. 1b). This is consistent with expectations, as (R_1, R_2) characterizes vowels. However, when f_0 exceeded this value of R_1 , the tuning line $R_1 = f_0$ was followed. This trend continued to 1 kHz for the vowels that do not use lip rounding ('hard', 'heard'), but for the vowels that use lip rounding ('hoard', 'who'd'), the data fell below the tuning line near 1 kHz. This may be because, with the lips rounded, it is uncomfortable or anatomically impossible to raise R_1 to 1 kHz.

This large shift in R_1 at high pitch means not only that vowels are shifted significantly on the vowel plane, but also that they will overlap considerably and converge to separations that are smaller than the characteristic length, λ , at which they become confused with one another⁶. This helps to explain the well-known difficulty in identifying words sung in the high range by sopranos^{8,9}, and may be one of the reasons why opera houses often use surtitles even for operas sung in the native language of their audience.

Elodie Joliveau, John Smith, Joe Wolfe

School of Physics, University of New South Wales, Sydney, New South Wales 2052, Australia
e-mail: john.smith@unsw.edu.au

1. Fant, G. *Speech Sounds and Features* (MIT, Cambridge, MA, 1973).
2. Sundberg, J. *Acustica* **32**, 89–96 (1975).
3. Sundberg, J. *The Science of the Singing Voice* (Northern Illinois Univ. Press, De Kalb, 1987).
4. Lindblom, B. E. F. & Sundberg, J. E. F. *J. Acoust. Soc. Am.* **50**, 1166–1179 (1971).
5. Sundberg, J. & Skoog, J. *J. Voice* **11**, 301–306 (1997).
6. Dowd, A., Smith, J. R. & Wolfe, J. *Lang. Speech* **41**, 1–20 (1997).
7. Epps, J., Smith, J. R. & Wolfe, J. *Measurement Sci. Technol.* **8**, 1112–1121 (1997).
8. Berlioz, H. *Grand Traité d'Instrumentation et d'Orchestration Modernes* (1844; transl. Clarke, M. C., Novello, London, 1882).
9. Hollien, H. et al. *J. Voice* **14**, 287–298 (2000).

Competing financial interests: declared none.

Dating the rise of atmospheric oxygen

A. Bekker¹, H. D. Holland¹, P.-L. Wang², D. Rumble III², H. J. Stein³, J. L. Hannah³, L. L. Coetzee⁴ & N. J. Beukes⁴

¹Department of Earth and Planetary Sciences, 20 Oxford Street, Harvard University, Cambridge, Massachusetts 02138, USA

²Geophysical Laboratory, Carnegie Institution of Washington, 5251 Broad Branch Road, Washington DC 20015, USA

³AIRIE Program, Department of Geosciences, Colorado State University, Fort Collins, Colorado 80523-1482, USA

⁴Department of Geology, Rand Afrikaans University, Auckland Park 2006, Johannesburg, South Africa

Several lines of geological and geochemical evidence indicate that the level of atmospheric oxygen was extremely low before 2.45 billion years (Gyr) ago, and that it had reached considerable levels by 2.22 Gyr ago. Here we present evidence that the rise of atmospheric oxygen had occurred by 2.32 Gyr ago. We found that syngenetic pyrite is present in organic-rich shales of the 2.32-Gyr-old Rooihoogte and Timeball Hill formations, South Africa. The range of the isotopic composition of sulphur in this pyrite is large and shows no evidence of mass-independent fractionation, indicating that atmospheric oxygen was present at significant levels (that is, greater than 10^{-5} times that of the present atmospheric level) during the deposition of these units. The presence of rounded pebbles of sideritic iron formation at the base of the Rooihoogte Formation and an extensive and thick ironstone layer consisting of haematitic pisolites and oolites in the upper Timeball Hill Formation indicate that atmospheric oxygen rose significantly, perhaps for the first time, during the deposition of the Rooihoogte and Timeball Hill formations. These units were deposited between what are probably the second and third of the three Palaeoproterozoic glacial events.

It was proposed long ago that the redox state of the atmosphere changed during the early Palaeoproterozoic¹. Sedimentary successions of age ≥ 2.45 Gyr include placer deposits that contain detrital uraninite, siderite and pyrite^{2,3}, reduced shallow-water facies of iron formations⁴, highly carbonaceous shales that are not enriched in redox-sensitive elements^{5,6} and palaeosols that are not oxidized⁶. Early diagenetic pyrite in these successions has $\delta^{34}\text{S}$ values consistent with a seawater sulphate content of $<200\ \mu\text{M}$ (ref. 7; see Table 1 for a definition of δ notation). In contrast, sedimentary successions younger than 2.22 Gyr contain red beds⁸, CaSO_4 -rich evaporites^{9,10}, and shallow-water iron formations that are oxidized¹¹. These successions overlie oxidized palaeosols¹² and have $\delta^{34}\text{S}$ records that are consistent with seawater sulphate concentrations $>200\ \mu\text{M}$ (ref. 13). The changes are strong evidence for a major increase in the O_2 content of the atmosphere between 2.45 and 2.22 Gyr ago, although it has been argued that they are related to differences in post-depositional alteration and to different tectonic settings rather than to an increase in the level of atmospheric oxygen^{14–17}.

The recent discovery of mass independent fractionation (MIF) in sulphur isotopes has provided a new tool for tracing changes in the oxygen content of the atmosphere¹⁸. Sulphur of sulphides and sulphates from sedimentary units older than 2.47 Gyr has values of MIF (expressed in terms of $\Delta^{33}\text{S}$; see Table 1 for a definition of Δ notation) ranging from -2.5‰ to $+8.1\text{‰}$, whereas all sulphides and sulphates younger than 1.9 Gyr have $\Delta^{33}\text{S}$ values $<|0.4\text{‰}|$ (refs 18–20). The only known mechanism for producing MIF in sulphur isotopes is photodissociation in the gas phase²¹, which has been observed in the modern atmosphere²². Preservation of large MIF signals in the Archaean record is probably related to the lack of an ozone shield in the atmosphere, allowing deep penetration of high energy ultraviolet and photochemical dissociation of SO_2 into elemental and water-soluble S species. The isotopic composition of elemental sulphur particles and water-soluble S species did not exchange completely in the Archaean atmosphere; hence some of their MIF signal was delivered to the Earth's surface²³. In an atmosphere with an oxygen content larger than $\sim 10^{-5}$ times the present atmospheric level (PAL)²³, sulphur species are oxidized to sulphate, exchange, and lose most of their MIF signal. The change from an anoxic to an oxygenated atmosphere therefore explains the absence of $\Delta^{33}\text{S}$ values $>0.4\text{‰}$ in the S isotopes of sulphides and sulphates during the past 1.9 Gyr, a period during which the

atmospheric partial pressure of oxygen (p_{O_2}) has been much greater than 10^{-5} PAL. These observations suggest that it should be possible to determine when p_{O_2} became $>10^{-5}$ PAL by defining the time when the range of MIF signals in sedimentary sulphides and sulphates decreased to $<0.4\text{‰}$. The results of our study of synsedimentary to early diagenetic pyrites in black shales of South Africa suggest that by 2.32 Gyr ago atmospheric p_{O_2} was already $>10^{-5}$ PAL.

Geological setting of studied units

The Rooihoogte Formation and the conformably overlying Timeball Hill Formation of the Pretoria Group, South Africa (Figs 1 and

Table 1 Sulphur isotope data and stratigraphic position of samples

Number	GC	$\delta^{34}\text{S}$ (‰)	$\delta^{33}\text{S}$ (‰)	$\delta^{34}\text{S}^*$ (‰)	$\delta^{33}\text{S}^*$ (‰)	$\Delta^{33}\text{S}_{\text{in}}$ (‰)	Depth (m)
Timeball Hill Formation							
EBA2/30-1	c	-25.56	-13.13	-25.89	-13.22	0.08	1,333.0–1,333.1
-2	c	-26.23	-13.51	-26.58	-13.60	0.04	1,333.0–1,333.1
-3	c	-25.6	-13.16	-25.93	-13.25	0.07	1,333.0–1,333.1
EBA2/34-1	d	-31.07	-16.18	-31.56	-16.31	-0.11	1,333.4–1,333.5
-2	c	-34.69	-17.93	-35.31	-18.09	0.03	1,333.4–1,333.5
-3	c	-33.28	-17.2	-33.85	-17.35	0.03	1,333.4–1,333.5
-4	d	-33.39	-17.12	-33.96	-17.27	0.17	1,333.4–1,333.5
-5	c	-33.45	-17.31	-34.02	-17.46	0.00	1,333.4–1,333.5
EBA2/38-1	d	-31.06	-16.18	-31.55	-16.31	-0.11	1,333.8–1,333.9
-2	d	-32.59	-17.02	-33.13	-17.17	-0.16	1,333.8–1,333.9
Rooihoogte Formation							
EBA2/67-1	d	-29.88	-15.59	-30.34	-15.71	-0.14	1,335.48–1,335.68
-2	c	-26.92	-13.78	-27.29	-13.88	0.13	1,335.48–1,335.68
-3	d	-23.91	-12.22	-24.20	-12.30	0.13	1,335.48–1,335.68
-4	d	-25.34	-12.82	-25.67	-12.90	0.27	1,335.48–1,335.68
EBA2/55-1	d	-29.07	-14.87	-29.50	-14.98	0.16	1,335.65–1,335.85
-2	d	-29.58	-15.29	-30.03	-15.41	0.01	1,335.65–1,335.85
EBA2/59-1	d	-25.76	-13.06	-26.10	-13.15	0.25	1,336.37–1,336.55
-2	d	-25.46	-12.83	-25.79	-12.91	0.33	1,336.37–1,336.55
EBA2/60-1	d	-28.92	-14.83	-29.35	-14.94	0.12	1,336.55–1,336.62
-2	c	-29.98	-15.45	-30.44	-15.57	0.06	1,336.55–1,336.62

In column GC (gas chromatograph record): c, clean (no significant peaks other than SF_6 were seen); d, dirty (peaks in addition to SF_6 were seen; the impurities were vented and did not enter the ion source of the mass spectrometer). Sulphur isotope ratios are represented by conventional δ notation with respect to CDT as $\delta^x\text{S} = 1,000[(^{x}\text{S}/^{32}\text{S})_{\text{sample}}/(^{x}\text{S}/^{32}\text{S})_{\text{CDT}} - 1]$, where x is either 33 or 34. $\Delta^{33}\text{S}$ is defined by the following equation: $\Delta^{33}\text{S} = \delta^{33}\text{S} - (0.5138\delta^{34}\text{S})$. The accuracy and precision of the technique as determined by repeated KrF excimer laser analyses of isotopically homogeneous sphalerite from Broken Hill, Australia, and comparison with CO_2 laser analyses of powders microdrilled adjacent to *in situ* analyses, is better than 0.2‰ (1 σ) for both $\delta^{33}\text{S}$ and $\delta^{34}\text{S}$, while values of $\Delta^{33}\text{S}$ are measured to $\pm 0.07\text{‰}$ (ref. 33). Several grains or different spots of the same grain were analysed in each sample; these are marked with a numerical extension after the sample number (for example, -1; -2; -3; and so on). Depths are given as m below surface.

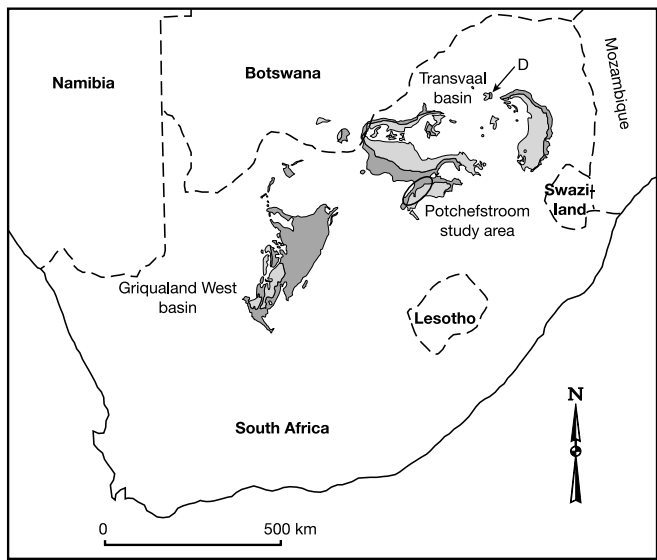


Figure 1 Map of Early Palaeoproterozoic sedimentary successions of South Africa. Distribution of the Chuniespoort (dark grey) and Pretoria (light grey) groups in the Transvaal structural basin and Ghaap (dark grey) and Postmasburg (light grey) groups in the Griqualand West structural basin of South Africa (after Coetzee²⁸). Both structural basins belong to the same depositional basin. 'D' indicates the location of the Duitschland Formation exposures.

2) overlie a prominent karstified unconformity on the Late Archaean Malmani carbonate platform and the $2,480 \pm 6$ -Myr-old Penge Iron Formation²⁴. Several lines of evidence suggest that the units below the unconformity are older than the rise of atmospheric oxygen: (1) early diagenetic pyrite of the Gamohaam Formation in the Campbellrand Subgroup carries a strong MIF signal¹⁸; (2) shallow-water iron formations of the Griqualand West basin,

which are slightly younger than the Penge iron formation, are reduced and lack a Ce anomaly^{4,25}; (3) $\delta^{34}\text{S}$ values of early diagenetic pyrites from these units cluster close to 0‰ and have a narrow range^{13,26}; (4) detrital uraninite and pyrite grains are present in the Black Reef Quartzite³. Unfortunately, the units immediately below the unconformity in the Transvaal basin (Penge Iron Formation and Tongwane Formation) lack pyrite that is suitable for MIF analysis.

On the basis of stacking patterns, the isopach map, facies analysis, and sedimentary structures, it follows that the Rooihoogte and Timeball Hill formations were deposited in a deltaic part of a basin open to the ocean on the southwest^{27–30}. Volcanic contributions to these units are lacking except in the extreme southwestern part of the basin, where the 30-m-thick Bushy Bend Lava occurs within the upper Rooihoogte Formation²⁸. The formations are sandwiched between two glacial diamictites, one at the base of the Rooihoogte Formation²⁸, the other at the top of the Timeball Hill Formation^{28,31}. We sampled a 3.6-m-thick section of highly carbonaceous shale that straddles the upper Rooihoogte and lower Timeball Hill formations in core EBA-2. The Rooihoogte Formation is condensed in this area, and the sampled section is only 12.5 m above the base of the formation. The units have been affected by lower greenschist facies metamorphism²⁸.

Description of the studied material

Pyrite in this section is confined to highly organic-rich layers. It occurs as nodules, mineralized microbial mats, disseminated grains, and laminated seams up to several centimetres in thickness. The nodules are up to several centimetres in diameter. They either have no internal structure or are composed of globules, concentrically laminated, fine-grained pyrite, or radial, bladed pyrite crystals. Where nodules with radial, bladed crystals (spherulites) are closely packed, they form a substrate for layers of fibrous pyrite crystals. The spherulites (Fig. 3) probably consisted initially of marcasite. The small pyrite globules have a diameter of 10–600 μm. The outer parts of the nodules are commonly overgrown by euhedral pyrite crystals. Laminae in the shale bend around nodules, suggesting that

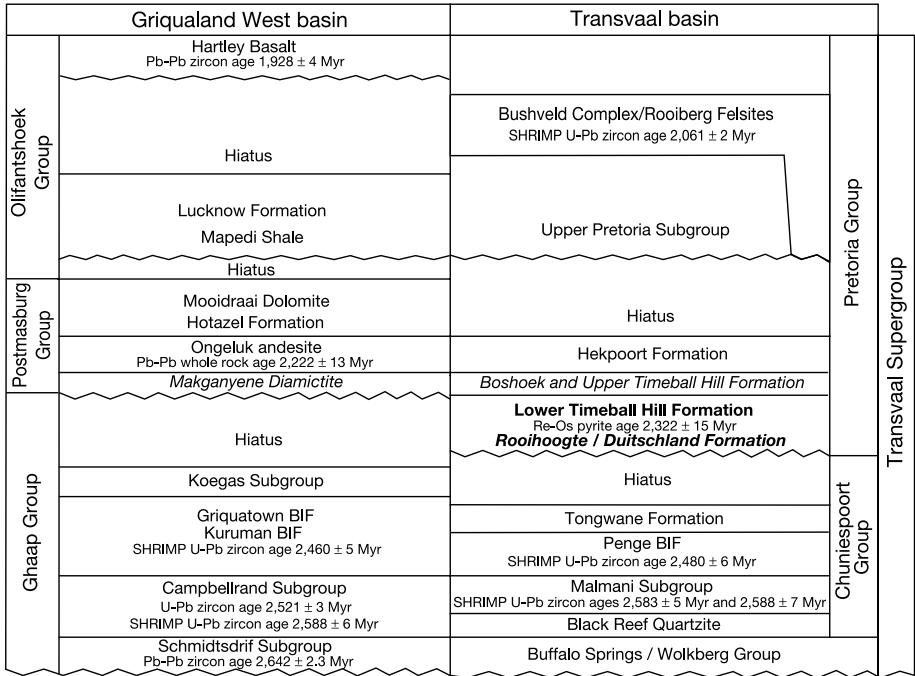


Figure 2 Correlation chart for the Transvaal Supergroup in the Griqualand West and Transvaal structural basins, South Africa (modified from ref. 39). Studied units are shown in bold; units in italics contain glacial diamictites. The Re-Os pyrite age is from ref. 38, the

age of the Kuruman iron formation is from ref. 44; see ref. 39 for references to other ages. BIF, banded iron formation; SHRIMP, sensitive high-resolution ion micro-probe.

the nodules predate compaction. Microbial mats and microbial mat rip-ups mineralized by pyrite form wavy-crinkly laminae which are similar to those in shales of the Mesoproterozoic Belt Group, Montana, and in many Phanerozoic units³². The laminated seams are fine-grained. Their parallel laminations are defined by variable crystal size and by variations in the concentration of clay and organic matter. Scanning electron microscopy (SEM) analyses indicate that the pyrite contains small amounts of sphalerite, chalcopyrite and galena. These sulphides also occur in veins that cut across bedding. Some of the small pyrite grains contain Se, Pb and, rarely, As. These textural and mineralogic characteristics as well as the association of the pyrite with organic-rich shales suggest an early diagenetic origin for all or nearly all of the pyrite.

Isotopic data

Drill core samples were cut parallel to bedding, polished, and then analysed. Early diagenetic pyrite grains larger than 0.5 mm in polished thick sections were pre-fluorinated, and SF₆ analyte was prepared by the *in situ* KrF excimer laser photoablation technique³³. The analyte was passed through a gas chromatograph, peaks from contaminants were vented, and analysis of the purified SF₆ was made on a ThermoFinnigan Mat-251 at the Geophysical Laboratory of the Carnegie Institution. On a $\delta^{33}\text{S}^*$ versus $\delta^{34}\text{S}^*$ diagram, the isotopic composition of sulphur in the Rooihoogte-Timeball Hill pyrites plots along or close to a line with slope 0.5114 that passes through the origin (Fig. 4; $\delta^{33}\text{S}^*$ and $\delta^{34}\text{S}^*$ are defined³⁴ as $\delta^x\text{S}^* = 1,000 \ln([\delta^x\text{S}/1,000] + 1)$, where x is 33 and 34, respectively). All of the data in Fig. 4 lie within (or are less than the analytical error from) the area defined by fractionation arrays with slopes ranging between 0.500 and 0.516, typical of mass-dependent fractionation processes³⁵. This indicates that the data in Fig. 4 can be explained without invoking any MIF.

Values of $\Delta^{33}\text{S}_{\text{in}}$ in Table 1 were calculated by means of the equation $\Delta^{33}\text{S}_{\text{in}} = \delta^{33}\text{S}^* - (0.5138 \delta^{34}\text{S}^*)$, where the slope 0.5138 is defined by more than 100 unpublished analyses done at the Geophysical Laboratory. All the values are $\leq 0.33\text{‰}$. This indicates that the range of the MIF signal in these data, if present at all, is much smaller than the range observed in most Archaean sulphides and sulphates but is similar to $\Delta^{33}\text{S}_{\text{in}}$ values in recent marine barite²⁰.

Implication for the rise of oxygen in the atmosphere

If the $\Delta^{33}\text{S}_{\text{in}}$ values of our data set are representative of sedimentary

sulphur during the deposition of the Timeball Hill and Rooihoogte formations, they are strong indications that the level of atmospheric O₂ was $>10^{-5}$ PAL. The lack of MIF $>0.33\text{‰}$ could be related to a hydrothermal source of sulphur that was not influenced by photodissociation in the atmosphere, or to local, complete mixing of S species affected by MIF. However, the highly negative $\delta^{34}\text{S}$ values of sulphur in the pyrite, consistently small $\Delta^{33}\text{S}$ values over the 3.5-m-thick section of organic-rich shales, and their deposition in a shallow-water epicontinental basin far from sites of contemporaneous volcanic activity, speaks for its origin via the bacterial reduction of normal seawater sulphate.

Two additional observations support the interpretation that the lack of (or small) MIF signal in the Rooihoogte and Timeball Hill pyrites is related to the rise of the atmospheric oxygen level before these units were deposited rather than to local, complete mixing of S species affected by MIF. First, the $\delta^{34}\text{S}$ values of the pyrites range from -34.7 to -23.9‰ . This range is consistent with the results of whole rock analyses in a previous study²⁶, suggesting that ^{34}S depletion in pyrites from these units is basin-wide. These values are much lower than those found in all older sedimentary units¹³. The small S isotope fractionation in the Archaean ocean has been explained by a seawater sulphate content of $<200 \mu\text{M}$ (ref. 7). The Timeball Hill and Rooihoogte S isotope data suggest a higher seawater sulphate content during their deposition, consistent with an intensification of the oxidative part of the S cycle rather than local hydrothermal influence. Second, sandstones overlying shales of the lower Timeball Hill Formation contain iron ore consisting of haematitic oolites and pisolites deposited in shallow-water fluvial and deltaic settings³⁶. These haematitic iron ores cover an area of $>100,000 \text{ km}^2$ and contain ore reserves of 6 billion tons at 40–55% Fe (refs 36, 37). Their presence implies that the level of atmospheric oxygen was at least high enough to oxidize Fe in shallow deltaic and fluvial settings. These data are all consistent with an oxygen content of the atmosphere $>10^{-5}$ PAL.

Until recently, the age of the Rooihoogte and Timeball Hill formations was only constrained to be between 2.48 and 2.22 Gyr (Fig. 2). A Re–Os age of $2,322 \pm 15 \text{ Myr}$ has been reported³⁸ for the early diagenetic pyrites described above. The rise of atmospheric oxygen to values $>10^{-5}$ PAL therefore began before $2,322 \pm 15 \text{ Myr}$ ago. The Deutschland Formation, which is preserved in the north-eastern part of the Transvaal basin (Fig. 1), is a time equivalent of the Rooihoogte Formation. It occupies the same stratigraphic position with respect to the prominent underlying karstified unconformity

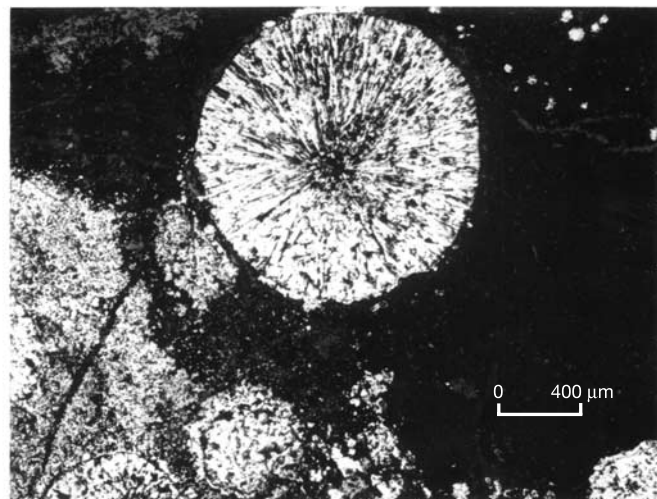


Figure 3 Reflected-light photomicrograph of pyrite spherulite in a polished thin section of sample EBA2/30-1.

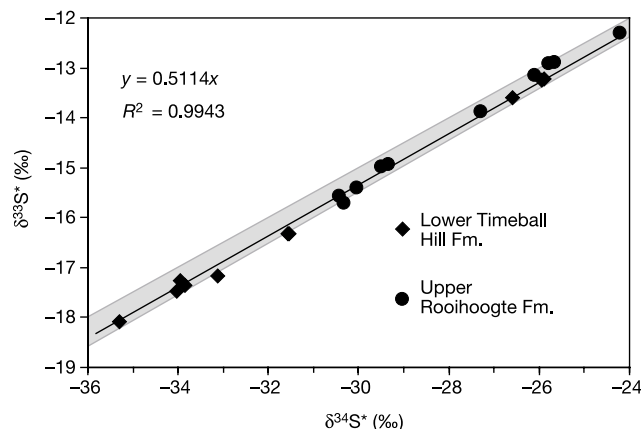


Figure 4 Plot of $\delta^{33}\text{S}^*$ versus $\delta^{34}\text{S}^*$ values in pyrite of the Rooihoogte and Timeball Hill formations. The grey area defines the field of isotopic data that can be produced by mass-dependent fractionation processes along the fractionation arrays with slopes ranging between 0.500 and 0.516 (ref. 35). Note that the $\delta^{33}\text{S}^*/\delta^{34}\text{S}^*$ ratio in the analysed samples lies on (or close to) the line of slope 0.5114.

surface and the overlying Timeball Hill Formation. There is a basal glacial diamictite in both the Deutschland and the Rooihoogte formations. Quartzites occur in the middle of both formations, and a chert breccia is present at their top²⁸. The upper part of the Deutschland Formation contains marine carbonates with $\delta^{13}\text{C}$ values as high as +10.1‰, suggesting the production of excess oxygen by biological carbon cycling³⁹. Their stratigraphic position is broadly correlative with that of the black shales of the Rooihoogte Formation, suggesting a possible link between the high relative burial rate of organic carbon and the rise of atmospheric oxygen.

There is still a time gap of ~150 Myr between the youngest sediments with a strong MIF signal (~2.47-Gyr-old Dales Gorge Member of the Brockman Iron Formation, Western Australia^{18,40}) and the oldest sediments with a small or no MIF signal (Rooihoogte Formation). The record of sedimentation in the Huronian Supergroup, Canada might fill this gap. The glacial diamictites of the Rooihoogte and Timeball Hill formations have recently been correlated on the basis of chemostratigraphy and event stratigraphy with the second and third of the three glacial diamictites of the Huronian Supergroup, Canada³⁹. Wing *et al.*⁴¹ found no $\Delta^{33}\text{S}$ values >0.3‰ in the sulphur isotopes of pyrite in shales that underlie and overlie the oldest glacial diamictite of the Huronian Supergroup in Ontario. This suggests that the loss of the MIF signal pre-dates the first glacial diamictite. However, the early diagenetic origin of the pyrite analysed by Wing *et al.* is not firmly established.

Although large MIF signals may have disappeared before the first glacial event, the O_2 content of the atmosphere seems to have remained very low until ~2,322 Myr ago. Detrital pyrite and uraninite grains occur in sandstones above the first glacial diamictite in North America⁴² and pebbles of siderite facies iron formation occur in a conglomerate at the base of the Rooihoogte Formation²⁸. The carbon isotope record, though still rather poorly defined for this time interval, does not indicate the presence of a positive excursion before the first glacial event or between the first and second glacial events⁴³. The first major increase in atmospheric oxygen may, therefore, have occurred during the deposition of the Rooihoogte and Timeball Hill formations. □

Received 5 August; accepted 28 November 2003; doi:10.1038/nature02260.

- MacGregor, A. M. The problem of the Precambrian atmosphere. *S. Afr. J. Sci.* **23**, 155–172 (1927).
- Rasmussen, B. & Buick, R. Redox state of the Archean atmosphere: Evidence from detrital heavy minerals in ca. 3250–2750 Ma sandstones from the Pilbara Craton, Australia. *Geology* **27**, 115–118 (1999).
- England, G. L., Rasmussen, B., Krapez, B. & Groves, D. I. Paleoenvironmental significance of rounded pyrite in siliclastic sequences of the Late Archean Witwatersrand Basin: Oxygen-deficient atmosphere or hydrothermal alteration? *Sedimentology* **49**, 1133–1156 (2002).
- Beukes, N. J. & Klein, C. Geochemistry and sedimentology of a facies transition—from microbanded to granular iron-formation—in the early Proterozoic Transvaal Supergroup, South Africa. *Precamb. Res.* **47**, 99–139 (1990).
- Yang, W. & Holland, H. D. The redox-sensitive trace elements, Mo, U, and Re in Precambrian carbonaceous shales: Indicators of the Great Oxidation Event. *Geol. Soc. Am. Abstr. Programs* **34**, 381 (2002).
- Holland, H. D. in *Early Life on Earth* (ed. Bengtson, S.) 237–244 (Columbia Univ. Press, New York, 1994).
- Habicht, K. S., Gade, M., Thamdrup, B., Berg, P. & Canfield, D. E. Calibration of sulfate levels in the Archean Ocean. *Science* **298**, 2372–2374 (2002).
- Chandler, F. W. Proterozoic redbed sequences of Canada. *Can. Geol. Surv. Bull.* **311** (1980).
- Chandler, F. W. Diagenesis of sabkha-related, sulphate nodules in the Early Proterozoic Gordon Lake Formation, Ontario, Canada. *Carbon. Evapor.* **3**, 75–94 (1988).
- El Tabakh, M., Grey, K., Pirajno, F. & Schreiber, B. C. Pseudomorphs after evaporitic minerals interbedded with 2.2 Ga stromatolites of the Yerriba basin, Western Australia: Origin and significance. *Geology* **27**, 871–874 (1999).
- Beukes, N. J. & Klein, C. in *Proterozoic Biosphere* (eds Schopf, J. W. & Klein, C.) 147–151 (Cambridge Univ. Press, Cambridge, UK, 1992).
- Rye, R. & Holland, H. D. Paleosols and the evolution of atmospheric oxygen: A critical review. *Am. J. Sci.* **298**, 621–672 (1998).
- Strauss, H. in *Precambrian Sedimentary Environments: A Modern Approach to Ancient Depositional Systems* (eds Altermann, W. & Corcoran, P. L.) 67–105 (Blackwell Science, Oxford, 2002).
- Dimroth, E. & Kimberley, M. M. Precambrian atmospheric oxygen: Evidence in the sedimentary distribution of carbon, sulfur, uranium, and iron. *Can. J. Earth Sci.* **13**, 1161–1185 (1976).

- Clemmey, H. & Badham, N. Oxygen in the Precambrian atmosphere: An evaluation of the geological evidence. *Geology* **10**, 141–146 (1982).
- Phillips, G. N., Myers, R. E. & Palmer, J. A. Problems with the placer model for Witwatersrand gold. *Geology* **15**, 1027–1030 (1987).
- Ohmoto, H. Evidence in pre-2.2 Ga paleosols for the early evolution of the atmospheric oxygen and terrestrial biota. *Geology* **24**, 1135–1138 (1996).
- Farquhar, J., Bao, H. & Thiemens, M. Atmospheric influence of Earth's earliest sulfur cycle. *Science* **289**, 756–758 (2000).
- Ono, S. *et al.* New insights into Archean sulfur cycle from mass-independent sulfur isotope records from the Hamersley Basin, Australia. *Earth Planet. Sci. Lett.* **213**, 15–30 (2003).
- Farquhar, J. *et al.* Mass-independent sulfur of inclusions in diamond and sulfur recycling on early Earth. *Science* **298**, 2369–2372 (2002).
- Farquhar, J., Savarino, J., Airieau, S. & Thiemens, M. H. Observation of wavelength-sensitive mass-independent sulfur isotope effects during SO_2 photolysis: Implications for the early atmosphere. *J. Geophys. Res.* **106**, 1–11 (2001).
- Romero, A. B. & Thiemens, M. Mass-independent sulfur isotopic compositions in sulfate aerosols and surface sulfates derived from atmospheric deposition: Possible sources of the MI anomaly and implications for atmospheric chemistry. *Eos* **83**(Fall Meet. Suppl.), B71A–0731 (2002).
- Pavlov, A. A. & Kasting, J. F. Mass-independent fractionation of sulfur isotopes in Archean sediments: Strong evidence for an anoxic Archean atmosphere. *Astrobiology* **2**, 27–41 (2002).
- Martini, J. E. J. The fluorite deposits in the Dolomite Series of the Marico District, Transvaal, South Africa. *Econ. Geol.* **71**, 625–635 (1976).
- Bau, M., Beukes, N. J. & Romer, R. L. Increase of oxygen in the Earth's atmosphere and hydrosphere between ~2.5 and ~2.4 Ga B.P. *Mineral. Mag.* **62**, 127–128 (1998).
- Cameron, E. M. Sulphate and sulphate reduction in early Precambrian oceans. *Nature* **296**, 145–148 (1982).
- Eriksson, K. A. The Timeball Hill Formation—A fossil delta. *J. Sedim. Res.* **43**, 1046–1053 (1973).
- Coetzee, L. L. *Genetic Stratigraphy of the Paleoproterozoic Pretoria Group in the Western Transvaal* Thesis, Rand Afrikaans Univ. (2001).
- Eriksson, P. G. & Reesko, B. F. F. Contourites associated with pelagic mudrocks and distal delta-fed turbidites in the Lower Proterozoic Timeball Hill Formation epicritic basin (Transvaal Supergroup), South Africa. *Sedim. Geol.* **120**, 319–335 (1998).
- Visser, J. N. J. The Timeball Hill Formation at Pretoria—A prograding shore-line deposit. *Ann. Geol. Surv. (S. Afr.)* **9**, 115–118 (1971) (1971–72).
- Visser, J. N. J. in *Earth's Pre-Pleistocene Glacial Record* (eds Hambrey, M. J. & Harland, W. B.) 180–184 (Cambridge Univ. Press, New York, 1981).
- Schieber, J. Microbial mats in terrigenous clastics: The challenge of identification in the rock record. *Palaio* **14**, 3–12 (1999).
- Hu, G., Rumble, D. & Wang, P.-L. An ultraviolet laser microprobe for *in situ* analysis of multi-sulfur isotopes and its use in measuring Archean sulfur isotope mass-independent anomalies. *Geochim. Cosmochim. Acta* **67**, 3101–3118 (2003).
- Miller, M. F. Isotopic fractionation and the quantification of ^{17}O anomalies in the oxygen three-isotope system: An appraisal and geochemical significance. *Geochim. Cosmochim. Acta* **66**, 1881–1889 (2002).
- Farquhar, J. *et al.* Multiple sulphur isotopic interpretations of biosynthetic pathways: Implications for biological signatures in the sulphur isotope record. *Geobiology* **1**, 27–36 (2003).
- Beukes, N. J., Dorland, H. C. & Gutzmer, J. Pisolitic ironstone and ferricrete in the 2.22–2.4 Ga Timeball Hill Formation, Transvaal Supergroup: Implications for the history of atmospheric oxygen. *Geol. Soc. Am. Abstr. Programs* **34**, 283 (2002).
- Schweiger, H. Genesis of the iron ores of the Pretoria Series, South Africa. *Econ. Geol.* **60**, 269–298 (1965).
- Hannah, J. L., Stein, H. J., Bekker, A., Markey, R. J. & Holland, H. D. Chondritic initial $^{187}\text{Os}/^{188}\text{Os}$ in Paleoproterozoic shale (seawater) and the onset of oxidative weathering. *Geochim. Cosmochim. Acta* **67** (2003) A-34.
- Bekker, A. *et al.* Chemostratigraphy of the Paleoproterozoic Deutschland Formation, South Africa: Implications for coupled climate change and carbon cycling. *Am. J. Sci.* **301**, 261–285 (2001).
- Mojzsis, S. J., Coath, C. D., Greenwood, J. P., McKeegan, K. D. & Harrison, T. M. Mass-independent isotope effects in Archean (2.5 to 3.8 Ga) sedimentary sulfides determined by ion microprobe analysis. *Geochim. Cosmochim. Acta* **67**, 1635–1658 (2003).
- Wing, B. A. *et al.* $\Delta^{33}\text{S}$, $\delta^{34}\text{S}$ and $\delta^{13}\text{C}$ constraints on the Paleoproterozoic atmosphere during the earliest Huronian glaciation. *Geochim. Cosmochim. Acta* **66**, A840 (2002).
- Roscoe, S. M. in *Geology of Canadian Mineral Deposit Types* (eds Eckstrand, O. R., Sinclair, W. D. & Thorpe, R. I.) P-1–10–23 (Geological Society of America, Denver, 1996).
- Bekker, A. *et al.* Response of the exosphere to the 2.48–2.45 Ga superplume event. *Geol. Soc. Am. Abstr. Programs* **32** (2000) A-135.
- Pickard, A. L. SHRIMP U-Pb zircon ages for the Palaeoproterozoic Kuruman Iron Formation, Northern Cape Province, South Africa: Evidence for simultaneous BIF deposition on Kaapvaal and Pilbara Cratons. *Precamb. Res.* **125**, 275–315 (2003).

Acknowledgements We thank D. Lange for help with the SEM work; W. Yang for help with sample preparation; and J. Brouwer for giving us access to drill core EBA-2, which was drilled by Gold Fields Ltd in the Potchefstroom area, and is now stored by the Geological Survey of South Africa. This study was supported by the NASA-Ames Research Center, NASA, PRF/ACS, NSF and South African Research Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.B. (abekker@fas.harvard.edu).

Notch activity acts as a sensor for extracellular calcium during vertebrate left–right determination

Ángel Raya^{1*}, Yasuhiko Kawakami^{1*}, Concepción Rodríguez-Esteban^{1*}, Marta Ibañes¹, Diego Rasskin-Gutman¹, Joaquín Rodríguez-León², Dirk Büscher¹, José A. Feijó^{2,3} & Juan Carlos Izpisua Belmonte¹

¹Gene Expression Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

²Instituto Gulbenkian de Ciencia, Rua Da Quinta Grande n.º 6, 2780-901 Oeiras, Portugal

³Centro de Biotecnologia Vegetal, Faculdade de Ciências, Universidade Lisboa, 1749-016 Lisboa, Portugal

* These authors contributed equally to this work

During vertebrate embryo development, the breaking of the initial bilateral symmetry is translated into asymmetric gene expression around the node and/or in the lateral plate mesoderm. The earliest conserved feature of this asymmetric gene expression cascade is the left-sided expression of *Nodal*, which depends on the activity of the Notch signalling pathway. Here we present a mathematical model describing the dynamics of the Notch signalling pathway during chick embryo gastrulation, which reveals a complex and highly robust genetic network that locally activates Notch on the left side of Hensen's node. We identify the source of the asymmetric activation of Notch as a transient accumulation of extracellular calcium, which in turn depends on left–right differences in H^+/K^+ -ATPase activity. Our results uncover a mechanism by which the Notch signalling pathway translates asymmetry in epigenetic factors into asymmetric gene expression around the node.

The disposition of most internal organs in vertebrates shows evident asymmetries along the left–right (LR) axis. The acquisition of proper LR asymmetries is achieved in a highly consistent manner during early embryo development. An evolutionarily conserved feature in the process of LR determination is the biased expression of *Nodal* in the left lateral plate mesoderm (LPM; reviewed in refs 1–5). Transgenic studies in mice have shown that left-sided expression of *Nodal* in the LPM depends on an earlier domain of *Nodal* expression around the node^{6,7}.

Notch activity plays an important role in the early phases of LR determination by directly regulating *Nodal* expression around the node in mouse^{8,9} and zebrafish⁹ embryos. The activity of the Notch signalling pathway depends on the complex interplay of several receptors, ligands and modulators (reviewed in ref. 10). In an effort to identify factors that regulate Notch activity around the node, we have characterized the dynamics of gene expression and interactions among relevant genes and proteins of this signalling pathway during early embryo gastrulation, and have generated a mathematical model to help us pinpoint those factors that are essential in LR determination.

LR asymmetry in components of the Notch pathway

We chose the chick embryo as the experimental model for these studies because of its unrivalled amenability for gene expression analysis during gastrulation, and also because *Nodal* expression around Hensen's node in the chick is overtly asymmetric during a sizeable window of time. We first investigated whether Notch activity is necessary for establishing proper LR asymmetry during chick embryo development. Indeed, blocking the Notch signalling pathway by overexpressing a dominant-negative form of the Notch pathway effector Suppressor of Hairless (Su(H)) resulted in laterality defects at both the morphological and the molecular level (see Supplementary Fig. 1), similar to those described for mouse embryos^{8,9}.

We then analysed the expression patterns of the known Notch genes (*Notch1–Notch3*) and ligands (*Delta-like 1 (Dll1)*, *Delta-like 3*,

Delta-like 4, *Serrate1* and *Serrate2*), as well as those of the Notch pathway modulators *Lunatic fringe (Lfng)* and *Radical fringe*, in chick embryos from Hamburger and Hamilton stage 3 (HH3) to HH8 (ref. 11) by *in situ* hybridization, and compared them with that of *Nodal*. The transcripts with a potential role in establishing LR asymmetry on the basis of their expression pattern are summarized in Fig. 1 (see also Supplementary Fig. 2). From these analyses it became evident that, before the appearance of the left-sided peri-nodal expression domain of *Nodal*, the Notch ligands *Dll1* and *Serrate1* show complementary patterns of expression that form a sharp anterior–posterior (AP) interface across Hensen's node (Fig. 1b, g, h).

Notably, *Dll1* transcripts at HH5 extended more anteriorly on the left side of the node in most embryos analysed (85%, $n = 125$; Fig. 1b, g). This asymmetry in *Dll1* expression is the earliest indication that Notch activity may be differentially lateralized. During HH3–HH7, *Lfng* is expressed in a complex, dynamic pattern of waves that sweep the AP axis of the embryo¹². We noticed that the fifth wave of *Lfng* is clearly asymmetric when it reaches the node at HH6: the medial-most part of the left stripe is anteriorly displaced with respect to the right (Fig. 1d). The location of the left-sided expression domain of *Nodal* at HH6 seems to coincide with the asymmetric region of *Dll1* expression previously established at HH5 (Fig. 1, compare e with b) and also overlaps with the anterior displacement of *Lfng* expression (Fig. 1, compare e with d).

Mathematical modelling of Notch activation

To gain a better understanding of the molecular mechanisms governing LR asymmetry determination and to be able to predict specific schemes that could be tested experimentally, the qualitative data collected from our gene expression analyses were used to model mathematically the dynamics of Notch activation during this process. The target of our model was to express *Nodal* differentially on the left side of the node while maintaining the hypothesis that the same mechanisms of signalling and interaction between cells operate in both the left and right peri-nodal regions.

We devised a network of interactions among messenger RNAs and proteins representing Notch1, Dll1, Serrate1 and Nodal (Fig. 2a). In brief, on the basis of the asymmetries detected in our gene expression analyses (Fig. 1) and the fact that Fringe has key roles in modulating Notch activity in several developmental contexts^{13–16}, we introduced *Lfng* as an external network modulator (see Fig. 2, Box 1 and Supplementary Methods). Furthermore, taking into account that LR asymmetries have been described to occur at stages earlier than HH4 (refs 1–4), we introduced a second external modulator acting along the LR axis in our model (see below).

Two assumptions in our theoretical model are that *Dll1* expression is induced by Notch activity, and that juxtaposition of *Dll1* and *Serrate1* expression domains strengthens Notch activity (Fig. 2a). To test the first assumption experimentally, we downregulated Notch signalling either by using DAPT (N-[N-(3,5-difluorophenacetyl-L-alanyl)]-S-phenylglycine *t*-butyl ester), a γ -secretase inhibitor known to block the Notch pathway *in vivo*¹⁷, or by overexpressing a dominant-negative version of Su(H). Either treatment resulted in a marked downregulation of *Dll1* expression (39% and 35%, $n = 80$ and $n = 95$, respectively; Fig. 3a–d).

To challenge experimentally the theoretical assumption that complementarities of *Dll1* and *Serrate1* expression domains are needed for proper Notch activity around Hensen's node, we removed Dll1 function by ectopically expressing a dominant-negative form of *Dll1* in the peri-nodal region. Embryos cultured in these conditions failed to activate left-sided expression of *Nodal* (38%, $n = 60$; Fig. 3e, f) and showed reversed heart looping (28%, $n = 48$; Fig. 3g). We also addressed the role of the juxtaposition of *Dll1* and *Serrate1* expression domains more directly by ubiquitously expressing *Dll1* in HH3–HH4 chick embryos. Notably, this manipulation resulted in alterations in LR patterning similar to those elicited by dominant-negative *Dll1* (87%, $n = 25$; Fig. 3h).

The positive role of *Lfng* as a modulator of Notch activity during LR determination was also tested experimentally. Our mathematical model considers that *Lfng* indirectly induces left-sided asymmet-

rical expression of *Dll1* at HH5 and directly activates the Notch pathway such that *Nodal* becomes expressed. This was experimentally confirmed by the retroviral delivery of *Lfng*, which caused ectopic right-sided expression of *Nodal* (48%, $n = 73$; Fig. 3i, j) and alterations in heart looping (34%, $n = 40$; Fig. 3k, l). Overall, our theoretical and experimental results underscore the importance of the juxtaposition of gene expression domains in biological design. Specifically, and during the establishment of LR asymmetries, the hyperactivation of Notch occurs precisely at the juxtaposition of *Dll1* and *Serrate1* expression and is strengthened where *Lfng* is expressed.

To identify potential modulators of Notch activity on the left or right side of Hensen's node, we mathematically explored the response generated by our model under different parameter values for specific initial conditions. Notably, our simulations show that *Dll1* and *Nodal* can be expressed asymmetrically by varying the binding rates of the complexes Notch1–Dll1 and Notch1–Serrate1. Thus, we find that the action of the LR modulator on binding rates is the most parsimonious way to reproduce the left and right expression patterns during stages HH4–HH6 (Fig. 2b, c).

In our model, *Lfng* provides a mechanism that ensures the appearance of *Nodal* by highly activating the Notch pathway. In a sense, this mechanism would operate as a synchronizer that would allow *Nodal* expression to be switched on at a very specific location and time (when the fifth *Lfng* wave sweeps the *Dll1*–*Serrate1* interface), rather than as an essential component of the system (Fig. 3m). This theoretical conclusion is in agreement with the absence of LR defects in mice that are mutant for *Lfng* (refs 14, 15; T. Gridley and R. Johnson, personal communications) and hints at the robustness of the system (understanding 'robustness' as constancy of outcome despite perturbations¹⁸).

In fact, one of the more noticeable properties of the genetic network characterized in our model is that it translates small variations in binding rates for the Notch–Dll1 and Notch–Serrate1 complexes into robust, side-specific expression of *Nodal*. The

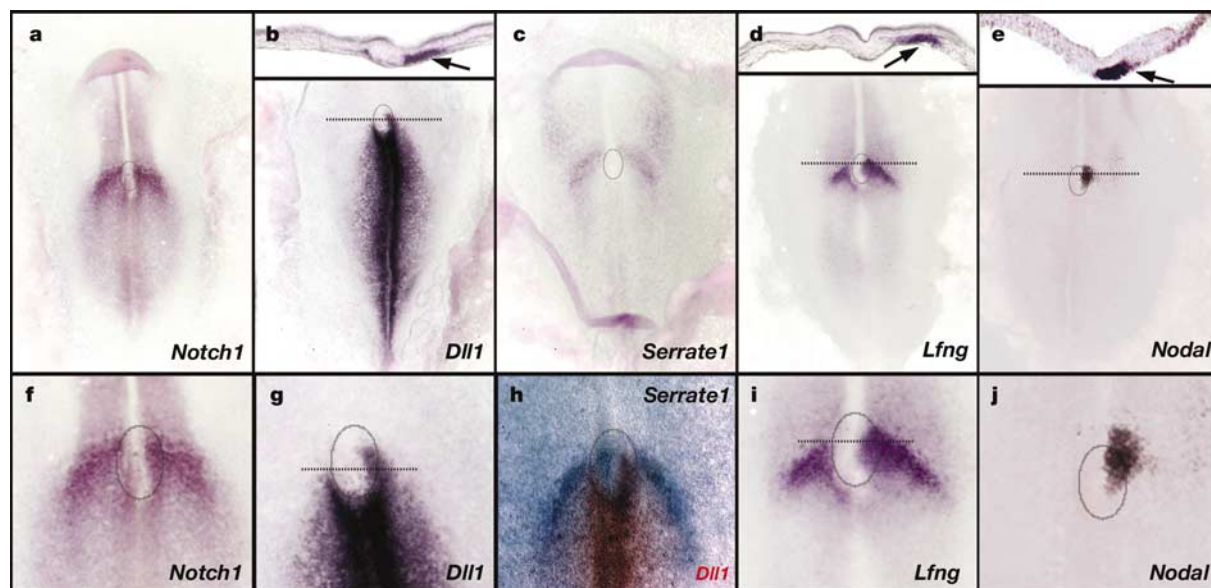


Figure 1 Asymmetries in the expression of Notch pathway components during chick gastrulation. **a**, *In situ* hybridization of HH6 chick embryos shows symmetrical expression of *Notch1* transcripts above Hensen's node. **b**, At HH5, *Dll1* transcripts become asymmetric in Hensen's node (dotted line in main panel). *Dll1*-positive cells are present on the left side of Hensen's node but are absent on the right side. In addition, *Dll1* transcripts extend anterior to the left side of Hensen's node (arrow in section shown on top). **c**, *Serrate1* transcripts at HH6 are observed in the form of two symmetrical stripes above and lateral to Hensen's node. **d**, At HH6, *Lfng* expression becomes asymmetric around

Hensen's node: the left stripe extends more medially and anteriorly than the right one (dotted line in main panel, arrow in section on top). **e**, *Nodal* transcripts appear at HH6, showing a clear signal on the left anterior side of Hensen's node (dotted line in main panel, arrow in section shown on top). **f**, **g**, **i**, **j**, Higher magnification of **a**, **b**, **d**, **e**, respectively, showing symmetric expression of *Notch1* and left-sided asymmetric expression of *Dll1*, *Lfng* and *Nodal*. **h**, Double-colour *in situ* hybridization for *Dll1* (red) and *Serrate1* (blue) in HH5 chick embryos, illustrating the juxtaposition of expression domains. All panels are ventral views, anterior is to the top. Ovals mark the position of Hensen's node.

system's ability to buffer perturbations is also reflected in our model, which tolerates not only the malfunction of *Lfng* (Fig. 3m) but also large fluctuations in several parameters (a detailed exploration of the sensitivity of the parameters is provided in Supplementary Methods), while allowing the normal left-sided expression of *Nodal*.

Notch activation depends on H^+/K^+ -ATPase activity

Our mathematical model predicts that a local domain of Notch hyperactivation reflects a pre-existing, transient LR asymmetry. Several factors and mechanisms are involved in the early events that break the initial bilateral symmetry in the embryo^{19–21}. Among them, the nodal flow created by monociliated cells in the mouse node^{22,23} seems to function in a parallel mechanism of LR determination that is unrelated to the one used by Notch^{8,9}. In addition, Sonic hedgehog (Shh), a signalling molecule that mediates left-

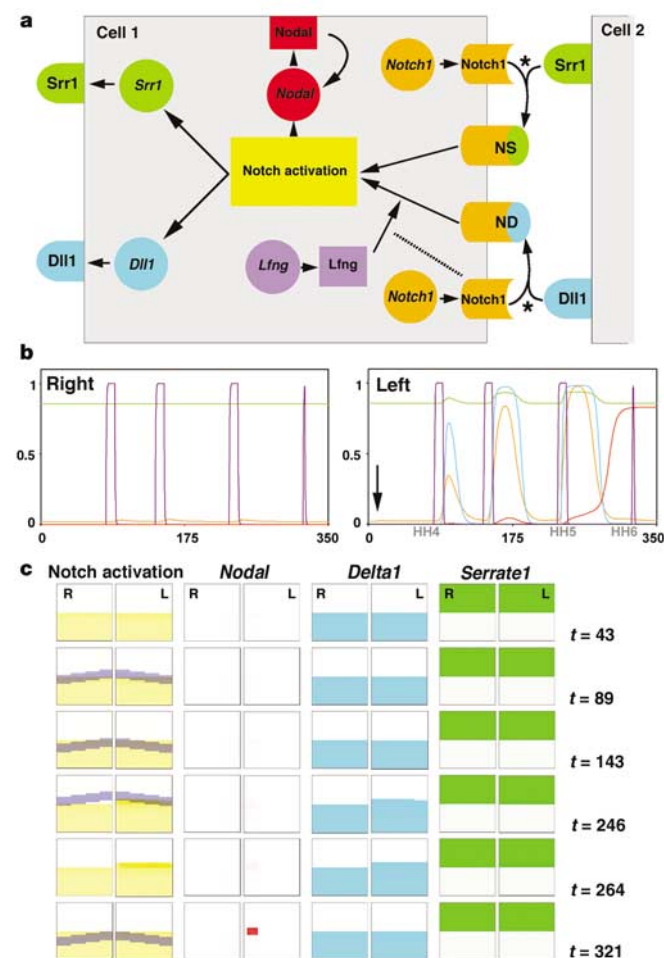


Figure 2 Mathematical model characterizing Notch activation dynamics. **a**, Network model. Circles symbolize mRNA, boxes represent proteins or protein complexes (ND, Notch1–Dll1; NS, Notch1–Serrate1). Two cells are shown. In cell 1, the full set of interactions can be seen. Asterisks denote the action of the external LR modulator in the formation of Notch complexes. The broken line indicates the regulation of Lfng action by Notch (see Box 1). **b**, Variation over time of Notch activation and *Nodal*, *Dll1* and *Serrate1* expression measured just above the *Dll1*–*Serrate1* interface (colour coding as in a). Arrow indicates the introduction of the external LR modulator. **c**, Simulation snapshots of gene expression and Notch activity. We consider the two lattices (left and right peri-nodal regions) to be initially equivalent. At time $t = 9$ min, the LR modulator is introduced on the left side. Later, four waves of *Lfng* sweep the left and right lattices, leading to transient asymmetric expression of *Dll1* more anteriorly and expression of *Nodal* only in the left peri-nodal region.

Box 1

A dynamic model of peri-nodal Notch activation

The mathematical model accounts for the formation of two complexes, Notch1–Dll1 (ND) and Notch1–Serrate1 (NS), which regulate a signalling pathway downstream of Notch1, called Notch activation (NA). We simplified this signal cascade into two steps: first, the receptor–ligand complexes activate the cascade; second, this cascade activates the mRNAs of *Dll1*, *Serrate1* and *Nodal* (refs 8, 9, 43, and Fig. 3). In addition, *Nodal* activates its own production via a positive feedback loop. The posterior regression of the node is also taken into account so that at stage HH4 the *Dll1*–*Serrate1* interface is just posterior to the node, whereas at stage HH6 it is anterior. *Lfng* appears dynamically as a wave that sweeps both lattices four times in a postero-anteriorly direction up to the most anterior part of the node, thereby influencing the activation of the Notch pathway cell autonomously^{12,44}. *Lfng* is known to encode a glycosyltransferase that acts on Notch protein^{45,46}. Several roles have been proposed for Fringe-mediated Notch signalling pathway modulation. We assume that Lfng modifies Notch so that its activity is enhanced on forming the complex with Delta.

A pair of independent two-dimensional lattices (left and right) of 16×32 cells is taken to characterize the peri-nodal left and right regions, respectively. In each lattice, every single cell is characterized by the amount of mRNA, protein and protein complexes, as well as by their mutual interaction (see Fig. 2a). The model has been formulated as a set of 13 coupled nonlinear delay and ordinary differential equations for each cell (see Supplementary Methods). The dynamics of each molecular species is a balance of its production (mRNA transcription and protein translation), its decay rate and the binding rate of complex formation. The interaction among macromolecular species is described according to mass-action kinetics. The pseudo steady-state approximation for activation rates is also taken into account.

Thus, if we consider that the dynamics of y are controlled by the activation through x and its own decay, the dynamic equations for y read:

$$\frac{dy}{dt} = \psi_k(x, h) - y; \psi_k(x, h) = \frac{x^k}{x^k + h^k}$$

where ψ_k is a saturating function with stiffness controlled by the Hill or cooperative coefficient (k , ref. 47), and parameter h is a measure of the threshold value above which activation is greater than 50%. By way of example, the equations for *Dll1* (*Dll*), Dll1 (*Dll*) and NA are written below:

$$\frac{dDl_i(t)}{dt} = a_d \psi_3(NA_i(t), h_d) - \mu_d Dl_i(t) \quad (1)$$

$$\frac{dDl_i(t)}{dt} = t_D Dl_i(t - T_D) - b_{ND} \sum_{j \in nn(i)} Dl_i(t) Notch_j(t) - \mu_D Dl_i(t) \quad (2)$$

$$\begin{aligned} \frac{dNA_i(t)}{dt} = & a_{NA} \psi_2(A[Lfng_i(t)/Notch_i(t)] \cdot ND_i(t) + B \cdot NS_i(t), h_{NA}) \\ & - \mu_{NA} NA_i(t) \end{aligned} \quad (3)$$

Equation (1) indicates that the activation of *Dll* in cell i depends on NA. If the Notch pathway is not active ($NA = 0$), then *Dll* is not expressed. As NA increases, *Dll* expression is enhanced until it saturates at a maximum stationary value of $Dl_{max} = \frac{a_d}{\mu_d}$, where a_d and μ_d are the transcription and decay rates respectively. As shown by the last term of equation (1), first order decay processes have been considered. According to equation (2), the change rate of Dll1 concentration depends on *Dll1* translation, Dll1 binding to Notch receptors in neighbouring cells and its decay rate. The delay term accounts for the time needed by large proteins to be translated, to be transported to the membrane and to become active (T_D). The communication between cells is established through juxtacrine (nearest neighbours, $nn(i)$) ligand–receptor interactions^{48–50}. In addition, equation (3) accounts for the activation of the Notch pathway through ND and/or NS. The ability of each complex to activate the Notch pathway is different; it is greater for ND, which is also modulated by Lfng.

sided expression of *Nodal* in the LPM²⁰, does not influence peri-nodal Notch activity⁹. It has been proposed that LR differences in H^+/K^+ -ATPase activity occur early during LR determination in chick embryos, resulting in gradients in membrane potential, and that this differential activity is crucial for proper LR establishment²⁴. These findings, however, do not explain how the initial symmetry is

broken in the first place²⁵, or how the differential H^+/K^+ -ATPase activity is translated into asymmetric gene expression around the node²⁶.

To investigate whether the Notch signalling pathway has a role in translating this differential activity, we analysed the appearance of the activation domain of Notch after inhibiting the H^+/K^+ -ATPase with omeprazole. Treatment of chick embryos with this compound results in bilateral expression of *Nodal* in the LPM (30%, $n = 18$; Fig. 4a, b; see also ref. 24). Notably, this treatment also induces a loss of asymmetric expression of *Dll1* (68%, $n = 52$; Fig. 4c, d) and *Lfng* (72%, $n = 40$; Fig. 4e, f) around Hensen's node. In addition, the expression of *Nodal* in the left peri-nodal region is greatly reduced after treatment with omeprazole (65%, $n = 30$; Fig. 4g, h). Despite this local decrease in *Nodal* transcripts, we could detect *Nodal* expression in the LPM, which at times was bilateral. We identified the origin of this ectopic *Nodal* expression in the fact that omeprazole treatment induces ectopic *Shh* expression (ref. 24 and Supplementary Fig. 3). These results indicate that the activity of the H^+/K^+ -ATPase may be at the base of the asymmetric event responsible for the appearance of Notch activation on the left side of Hensen's node.

LR asymmetry in extracellular Ca^{2+}

A differential H^+/K^+ -ATPase activity across Hensen's node creates a gradient in membrane potentials, which in turn is likely to originate a differential ion flux across the LR axis²⁴. It has been proposed that molecules that are small enough to pass through gap junctions and that are charged under physiological conditions will be accumulated on one side of the embryo's node; the proposed candidates include Ca^{2+} , inositol phosphates, cyclic nucleotides and neurotransmit-

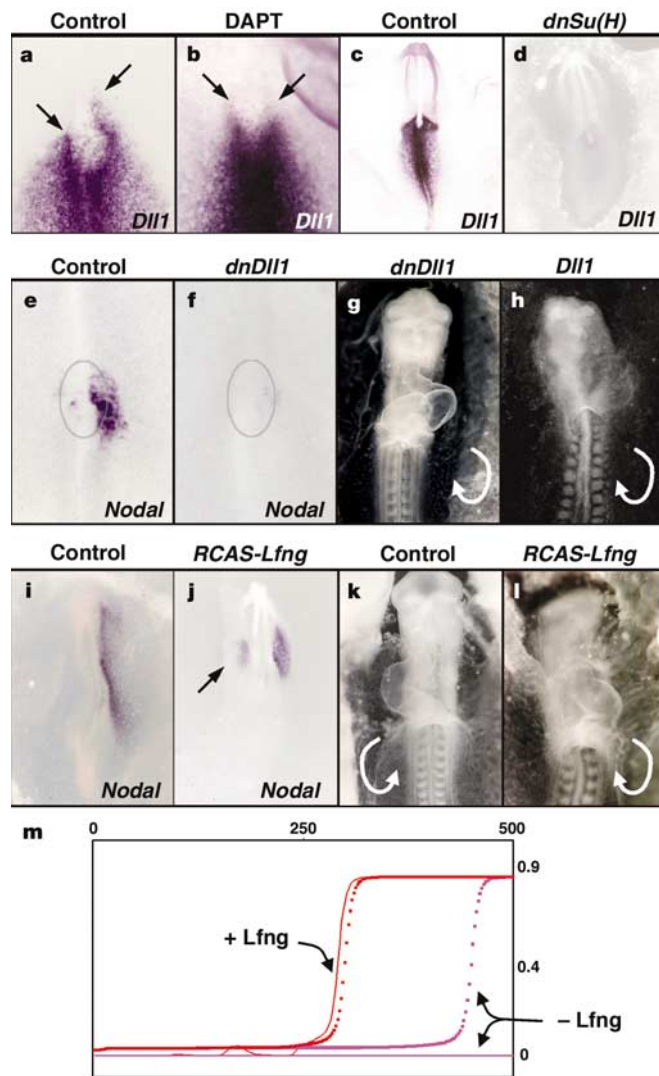


Figure 3 Experimental test of the theoretical assumptions and predictions.

a, b, Downregulation of Notch signalling by treatment with the γ -secretase inhibitor DAPT resulted in loss of the left asymmetric domain of *Dll1* expression (**b**, compare with control in **a**). Arrows point to anterior expression around Hensen's node. **c, d**, Overexpression of a dominant-negative form of the Notch effector Su(H) [*dnSu(H)*] downregulated *Dll1* expression in HH8 embryos (**d**, compare with control in **c**). **e–h**, Disruption of the *Dll1*–*Serrate1* expression interface by overexpressing a dominant-negative form of *Dll1* (*dnDll1*) resulted in a failure to activate left-sided *Nodal* expression (**f**, compare with control in **e**) and reversed heart looping (**g**). Similar phenotypic alterations in LR patterning were obtained after ubiquitous expression of *Dll1* (**h**). **i–l**, Compared with control embryos (**i, k**), ectopic expression of *Lfng* activated ectopic *Nodal* expression in the right LPM (arrow in **j**) and resulted in alterations of the normal rightward looping of the heart tube (**l**). All embryo images are ventral views, anterior is to the top. Ovals mark the position of Hensen's node. Semicircular arrows indicate direction of heart looping. **m**, Numerical results of our theoretical model for the temporal evolution of *Nodal* expression in the left peri-nodal region in normal conditions (red lines) and in conditions simulating *Lfng* ablation (pink lines). Unbroken lines are registered in a cell in the juxtaposed expression domain; broken lines correspond to cells in the *Dll1* domain exactly at the border with *Serrate1*. *Lfng* action results in a broader and earlier domain of *Nodal* expression.

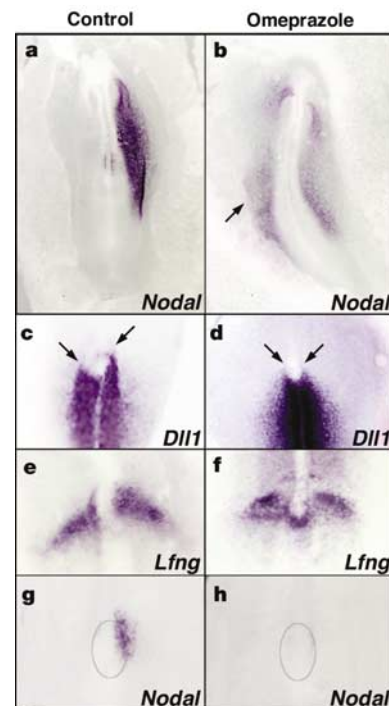


Figure 4 Differential H^+/K^+ -ATPase activity regulates Notch signalling during LR determination. **a, b**, Treatment of HH3–HH4 chick embryos with omeprazole, a pharmacological inhibitor of H^+/K^+ -ATPase, caused ectopic expression in the right LPM of *Nodal* (**b**, arrow). **c–f**, Treatment of chick embryos with omeprazole also prevented the appearance of the left-sided asymmetric expression domain of *Dll1* (**c, d**) and *Lfng* (**e, f**). Arrows point to anterior expression around Hensen's node. **g, h**, Analysis of *Nodal* expression after omeprazole treatment shows a downregulation of *Nodal* transcripts in the peri-nodal region (**h**). All panels are ventral views, anterior is to the top. Ovals mark the position of Hensen's node.

ters²⁷. Because our theoretical model underscores the crucial role of an external LR modulator that modifies receptor–ligand binding affinities, we reasoned that accumulation of Ca^{2+} on one side of the peri-nodal region could locally modulate the activity of the Notch signalling pathway. In this respect, structural analyses of epidermal growth factor (EGF)-like domain-mediated protein–protein interactions suggest that the receptor–ligand affinity and the activity of the Notch signalling pathway are extremely sensitive to Ca^{2+} concentrations^{28,29}.

We therefore sought to determine whether the extracellular distribution of Ca^{2+} presented any LR bias during chick gastrulation. The use of low-affinity, cell-impermeant Ca^{2+} indicators allowed us to visualize by two-photon excitation microscopy the existence of transient, local accumulations of extracellular Ca^{2+} on the left side of Hensen's node from HH4 to HH6 (Fig. 5a–c). These differences in extracellular Ca^{2+} concentration on either side of the node were no longer evident at HH8 (Fig. 5d). During the revision of this manuscript, evidence for an LR asymmetric distribution of intracellular Ca^{2+} concentrations around the mouse node was reported³⁰. The relationship, if any, between this asymmetry in cytoplasmic Ca^{2+} concentrations and the asymmetry reported here warrants further investigation.

To gain further insight into the mechanism responsible for LR

differences in extracellular Ca^{2+} concentrations, we cultured HH4–HH6 chick embryos in the presence of omeprazole and analysed the effects of this treatment on Ca^{2+} . A brief (5-min) incubation with omeprazole resulted in the disappearance of the LR difference in Ca^{2+} concentrations (100%, $n = 30$; Fig. 5e–i), showing that the gradient in extracellular Ca^{2+} is a consequence of the differential H^+/K^+ -ATPase activity. These results suggest that the alterations induced by omeprazole on LR determination and on the generation of the Notch activation domain may be related, at least in part, to the suppression of a gradient in extracellular Ca^{2+} .

To explore this hypothesis further, we investigated whether restoring asymmetric Ca^{2+} concentrations on the left side of the embryo could rescue the alterations induced by inhibiting the asymmetric H^+/K^+ -ATPase activity. Implantation of beads soaked in omeprazole into HH3–HH4 chick embryos resulted in reversed heart looping in a large proportion of embryos (25%, $n = 130$; Fig. 5j–k). Notably, implantation of Ca^{2+} -saturated agarose plugs on the left peri-nodal region of HH4–HH5 chick embryos completely prevented the LR patterning alterations induced by omeprazole (100%, $n = 14$; Fig. 5l).

Ca^{2+} modulates Notch and determines LR asymmetry

The role of the asymmetric concentrations of extracellular Ca^{2+} on

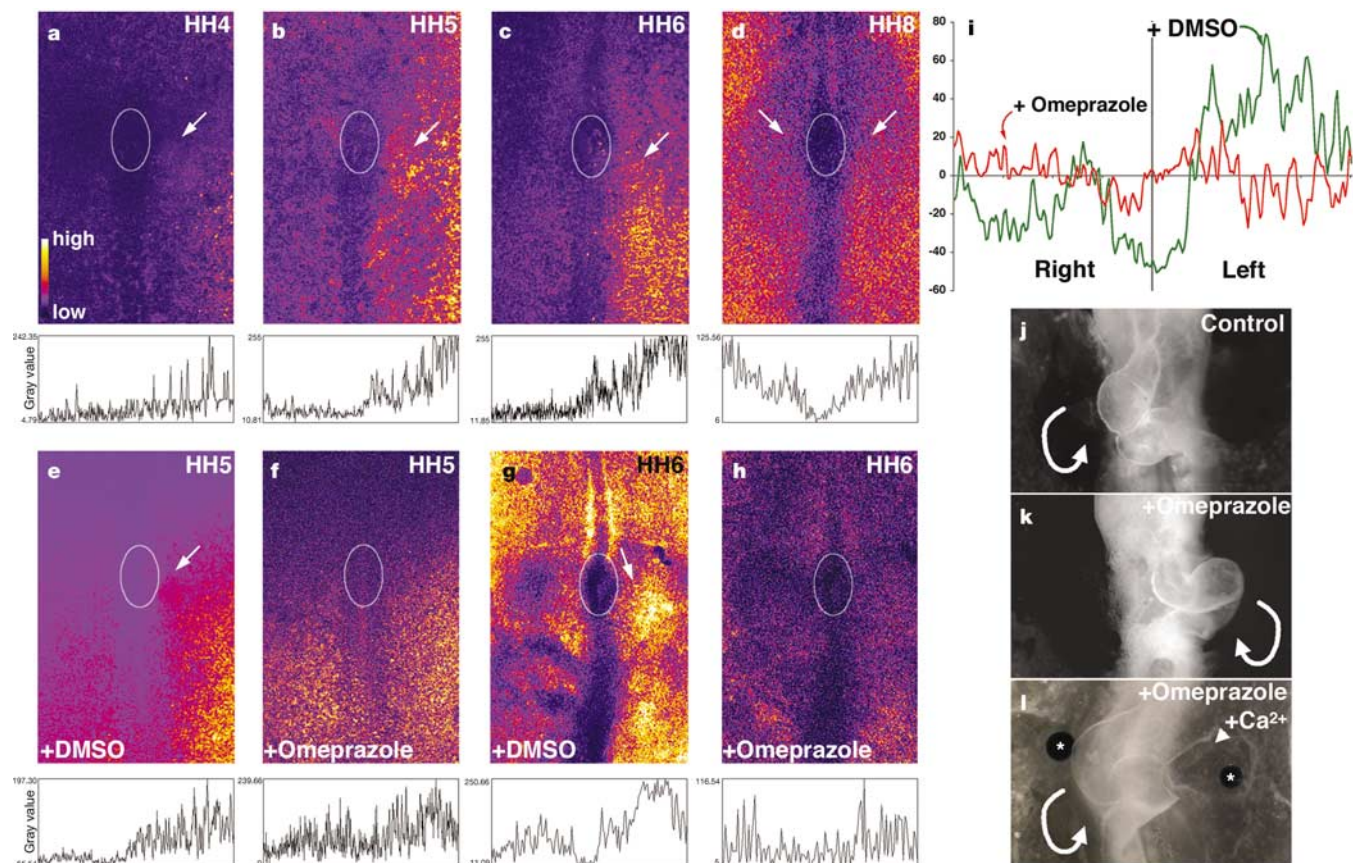


Figure 5 Asymmetric domains of increased Ca^{2+} during LR asymmetry determination. **a–d**, Extracellular Ca^{2+} during chick gastrulation was visualized with a cell-impermeant version of the Ca^{2+} indicator Calcium Green-5N. Asymmetric localized domains of increased Ca^{2+} (arrows) appeared on the left side of Hensen's node during HH4–HH6 and became symmetric at HH8. **e–h**, Two-photon excitation microscopy images of two representative experiments using chick embryos at HH5 (**e**, **f**) and HH6 (**g**, **h**) show that incubation with omeprazole (**f**, **h**) resulted in disappearance of the asymmetric Ca^{2+} distribution, which was previously visualized in the same embryos treated with the vehicle (DMSO) alone (arrows in **e**, **g**). Embryo views are ventral, anterior is to the top. Ovals mark the position of Hensen's node. In **a–h**, grey levels representing fluorescence intensities

have been pseudo-coloured with a purple to white scale, and the line scans at the bottom of the images were traced according to Hensen's node coordinates to quantify differences in fluorescence in the same embryos. **i**, Graphs representing the average of normalized fluorescence intensity values for four independent experiments in which Ca^{2+} concentrations were analysed after consecutive incubation with DMSO or omeprazole. **j–l**, Alterations in LR organ asymmetries induced by omeprazole (**k**) were prevented by raising Ca^{2+} around the left side of the node (**l**). Semicircular arrows indicate the direction of heart looping. The beads soaked in omeprazole (asterisks) and the Ca^{2+} -saturated agarose plug (arrowhead) are visible in **l**.

LR patterning was directly tested by analysing molecular and morphological landmarks of LR asymmetry after manipulating the extracellular Ca^{2+} concentrations by applying beads coated with the Ca^{2+} -specific chelator BAPTA to HH3 chick embryos. Treatment with BAPTA beads resulted in a high occurrence of the reversal of heart looping (44%, $n = 50$; Fig. 6a, b), an absence of *Nodal* expression around Hensen's node and subsequently in the LPM (35%, $n = 35$; Fig. 6c, d), and a failure to increase activation of Notch signalling around the node, as evaluated by the symmetry of *Dll1* expression at HH5 (33%, $n = 40$; Fig. 6f, g) and of *Lfng* expression at HH6 (40%, $n = 10$; Fig. 6h, i). These results indicate that activation of Notch on the left side of Hensen's node may depend on the local accumulation of extracellular Ca^{2+} . This was

further confirmed by expressing active Notch (the intracellular domain of Notch) on the left side of HH3 chick embryos, which rescued the LR defects induced by clamping extracellular Ca^{2+} concentrations with BAPTA (82%, $n = 11$; Fig. 6e).

To investigate whether the mechanism by which extracellular Ca^{2+} regulates Notch activity is direct, we used a co-culture system in which Notch-responding cells are activated by direct contact with cells expressing either Delta1 or Serrate1 ligands³¹ and Notch activity is measured by means of a *Hes1*-luciferase reporter³². The use of a *Nodal*-based reporter was precluded by the fact that Notch-mediated activation of *Nodal* is restricted to a specific cellular and developmental context (that is, the peri-nodal region of the gastrulating embryo^{8,9}), and cannot be mimicked *in vitro* in a co-

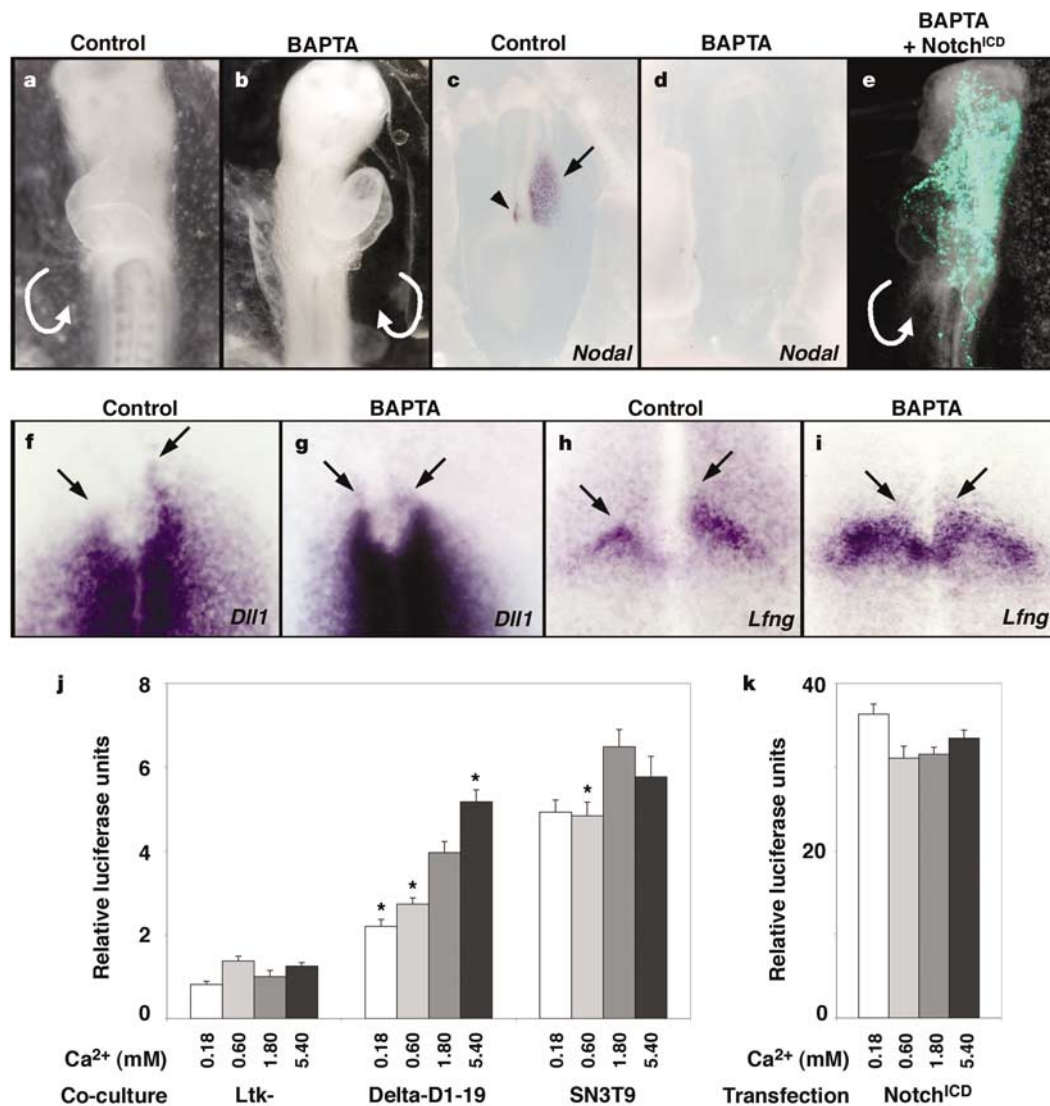


Figure 6 Extracellular Ca^{2+} concentrations directly regulate Notch activity and determine LR asymmetry. **a–i**, Clamping extracellular Ca^{2+} concentrations by applying the Ca^{2+} -specific chelator BAPTA coupled to polystyrene beads resulted in alterations in heart looping (**b**), a failure to induce left-sided *Nodal* expression in the node (arrowhead in **c**) and LPM (arrow in **c**), and a loss of the asymmetric expression of *Dll1* (**g**) and *Lfng* (**i**). Arrows in **f–i** point to the peri-nodal expression. **e**, Electroporation of an active form of Notch (Notch^{ICD}) on the left side of the embryo prevented the alterations in LR organ asymmetry induced by BAPTA. A plasmid expressing EGFP was co-injected to confirm left-side-specific electroporation. Shown is the superposition of dark-field and fluorescence images. Embryo views are ventral, anterior is to the top. Semicircular arrows mark the direction of heart looping. **j**, Notch activity was assayed by co-culturing Notch-

responding cells transiently transfected with a *Hes1*-luciferase reporter construct with the mouse fibroblast parental cell line (Ltk-) or with stable lines expressing *Dll1* (Delta-D1-19) or *Jagged1* (SN3T9) in culture medium containing the indicated concentrations of Ca^{2+} . Asterisks indicate statistically significant ($P < 0.05$) differences when compared with Notch activity elicited in medium containing 1.8 mM Ca^{2+} (two-tail Student's *t*-test). **k**, Notch activity induced independently of ligand by the co-transfection of active Notch (Notch^{ICD}) did not depend on extracellular Ca^{2+} in the concentration range tested. Luciferase activity was measured after 12 h of co-culture. β -Galactosidase activity was used to normalize for transfection efficiency. Values are expressed relative to the basal Notch activity of Ltk- co-cultures in culture medium containing 1.8 mM Ca^{2+} . Error bars represent the s.d. of three experiments with triplicate measurements.

culture assay (data not shown). By using an approach similar to the one used here, it has been shown that the complete chelation of Ca^{2+} from culture medium results in a ligand-independent induction of Notch activity, owing to dissociation of the non-covalent bond that holds together the intracellular and transmembrane domains of Notch³³. For this reason, we avoided the use of Ca^{2+} chelators and non-physiologically low concentrations of extracellular Ca^{2+} .

To examine the effect of extracellular Ca^{2+} concentration in Notch activity induced by Delta1 or Serrate1 binding, we used a Ca^{2+} -free medium supplemented with known amounts of CaCl_2 . Using this system, we were able to detect activation of Notch signalling in a Ca^{2+} - and ligand-dependent manner (Fig. 6j). Notably, we detected statistically significant differences between Notch activities elicited by Delta1 in the presence of Ca^{2+} levels three times lower or higher than the physiological extracellular Ca^{2+} concentration (1.8 mM).

To confirm that the modulation of Notch activity by extracellular Ca^{2+} actually depends on ligand–receptor interactions, we examined the effect of varying Ca^{2+} concentration on the activity of the *Hes1*-luciferase reporter induced by active Notch. We observed potent activation of the Notch reporter that was independent of the Ca^{2+} concentration in the culture medium (Fig. 6k). These results provide strong experimental evidence for a direct mechanism by which extracellular Ca^{2+} concentrations are sensed by the Notch signalling pathway.

Discussion

Our findings represent, to our knowledge, the first evidence that connects extracellular Ca^{2+} signalling with the determination of LR organ asymmetries. Although free cytosolic Ca^{2+} signalling has been described to occur and to be instrumental throughout embryonic development³⁴, the dynamics of extracellular Ca^{2+} concentrations has been largely overlooked. Our experiments not only identify the existence of localized domains of increased Ca^{2+} concentrations in the gastrulating chick embryo, but also provide direct evidence for a mechanism by which extracellular Ca^{2+} is sensed by the Notch signalling pathway and translated into differential gene expression.

The role of the Notch signalling pathway in providing polarity cues in various organisms, from nematodes to mammals, is well established. Our studies underscore the central role of this pathway during LR determination and identify the existence of a transient, asymmetric domain of Notch activation in the left peri-nodal region. The appearance of this domain depends on a spatio-temporal concurrence of both genetic and epigenetic factors, such that genetic information arranged along the AP axis merges with LR information relayed by Ca^{2+} movements to generate a potent amplification system. The participation of the Notch pathway in LR determination can be understood as a transient and robust mechanism that, acting as a sensor, amplifies local genetic and electro-chemical fluctuations to generate directly a domain of *Nodal* expression. Mechanisms such as the one reported here for organ asymmetry could be operating in other developmental or physiological contexts for the translation of transient epigenetic information into stable patterns of gene expression and, thus, heritable cell fates. □

Methods

Retroviral infection and DNA electroporation

Chick embryos were explanted and grown *in vitro* as described³⁵. RCAS BP(A) retroviral stocks containing dominant-negative *Su(H)*, dominant-negative *Dll1* or full-length *Lfng* were produced as described³⁶. We infected embryos by blastoderm injection at stages HH2–HH4. For electroporation of rat *Delta1* or active *Notch* (both cDNAs were gifts from C. Kintner, Salk Institute, La Jolla, CA), the full-length coding regions were subcloned into a modified pCAGGS vector (Y.K., unpublished data), and a 0.3–0.4 mg ml^{−1} solution of the vector was injected at HH2–HH3. DNA electroporation was done essentially as described³⁷. We used five square 50-ms pulses of 10 mV, spaced 350 ms apart. The

efficiency and extension of electroporation was monitored by co-injecting 0.05 mg ml^{−1} of pCAGGS–EGFP with the test DNA and examining the fluorescence of enhanced green fluorescent protein (EGFP).

Bead implantation and *in situ* hybridization

BAPTA coupled to polystyrene beads (Calcium Sponge, Molecular Probes) or AffiGel blue beads (BioRad) soaked in DAPT (γ -secretase inhibitor IX, Calbiochem) or omeprazole (Sigma), or small plugs of 25 mM CaCl_2 in 0.5% agarose were applied at stages HH3–HH4 on either side of the chick blastoderm near Hensen's node. Embryos were left to develop and processed for *in situ* hybridization³⁸ or morphological analysis. Sectioning of stained embryos was done as described³⁹. We carried out double-colour *in situ* hybridization essentially as described⁴⁰. Detailed descriptions of the RNA probes and constructs used are available from the authors on request.

Cell culture and reporter assays

The human cervical cancer cell line C-33A (ATCC HTB-31), which expresses *Notch1* (ref. 41) and responds to Notch signalling on the binding of Delta or Serrate ligands (C. Fryers and K. Jones, personal communication), was used as a reporter line in assays of Notch activity. Cells were plated on gelatine-coated 48-well clusters and transfected with different combinations of a *Hes1*-luciferase reporter (ref. 32; a gift from C. Kintner), active Notch and CMX- β -gal plasmids by using standard calcium phosphate transfection protocols (CellPhect, Amersham Biosciences). After 24 h, these cells were co-cultured with either the parental Ltk- mouse fibroblast cell line (ATCC CCL-1.3) or stable derivatives expressing *Delta1* or *Jagged1* (D1-19 (ref. 31) or SN3T9 (ref. 42), respectively; both lines were gifts from C. Fryers and K. Jones, Salk Institute) in culture medium containing different concentrations of CaCl_2 . Co-cultures were stopped at 2–12 h and assayed for luciferase activity by standard methods. We used β -galactosidase activity to normalize for transfection efficiency. The activity of the reporter is expressed relative to basal values obtained in co-cultures with Ltk- cells in medium containing 1.8 mM CaCl_2 .

Mathematical modelling

Details of the modelling are given in the Supplementary Methods.

Extracellular Ca^{2+} imaging

Embryos at stages HH3–HH6 were placed in a plastic ring as described³⁵. The plastic ring with the embryo was placed in a Wilco dish with a glass bottom and incubated at 38 °C for 15 min with 10 μ l of 40 μ M Calcium Green-5N, hexapotassium salt (Molecular Probes) diluted in Ca^{2+} -free PBS, pH 7.4. Embryos were imaged on either a 1024 MP (BioRad) or a Radiance 2100 (BioRad) instrument using $\times 10$ –20 PlanFluo lenses (Nikon). Two-photon excitation was done at 820 nm with Coherent Mira-Verdi or Tsunami lasers using a pulse width shorter than 150 fs and a 522/DF35 emission filter. Drug-treated embryos were incubated in the presence of 5 μ l of 7 μ M omeprazole (Sigma), diluted in Ca^{2+} -free PBS from a stock in dimethyl sulphoxide (DMSO; final DMSO concentration <0.5%), and were imaged after omeprazole addition. Before the addition of omeprazole, the embryos were imaged after incubation with equivalent concentrations of DMSO. Images were processed with the ImageJ software (NIH).

Received 11 August; accepted 23 October 2003; doi:10.1038/nature02190.

- Burdine, R. D. & Schier, A. F. Conserved and divergent mechanisms in left–right axis formation. *Genes Dev.* **14**, 763–776 (2000).
- Capdevila, J., Vogan, K. J., Tabin, C. J. & Izpisua Belmonte, J. C. Mechanisms of left–right determination in vertebrates. *Cell* **101**, 9–21 (2000).
- Mercola, M. & Levin, M. Left–right asymmetry determination in vertebrates. *Annu. Rev. Cell Dev. Biol.* **17**, 779–805 (2001).
- Hamada, H., Meno, C., Watanabe, D. & Saijoh, Y. Establishment of vertebrate left–right asymmetry. *Nature Rev. Genet.* **3**, 103–113 (2002).
- Bisgrove, B. W., Morelli, S. H. & Yost, H. J. Genetics of human laterality disorders: insights from vertebrate model systems. *Annu. Rev. Genomics Hum. Genet.* **4**, 1–32 (2003).
- Brennan, J., Norris, D. P. & Robertson, E. J. Nodal activity in the node governs left–right asymmetry. *Genes Dev.* **16**, 2339–2344 (2002).
- Saijoh, Y., Oki, S., Ohishi, S. & Hamada, H. Left–right patterning of the mouse lateral plate requires nodal produced in the node. *Dev. Biol.* **256**, 160–172 (2003).
- Krebs, L. T. et al. Notch signaling regulates left–right asymmetry determination by inducing Nodal expression. *Genes Dev.* **17**, 1207–1212 (2003).
- Raya, A. et al. Notch activity induces Nodal expression and mediates the establishment of left–right asymmetry in vertebrate embryos. *Genes Dev.* **17**, 1213–1218 (2003).
- Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. Notch signaling: cell fate control and signal integration in development. *Science* **284**, 770–776 (1999).
- Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *J. Morphol.* **88**, 49–92 (1951).
- Jouve, C., Iimura, T. & Pourquie, O. Onset of the segmentation clock in the chick embryo: evidence for oscillations in the somite precursors in the primitive streak. *Development* **129**, 1107–1117 (2002).
- Irvine, K. D. Fringe, Notch, and making developmental boundaries. *Curr. Opin. Genet. Dev.* **9**, 434–441 (1999).
- Evrard, Y. A., Lun, Y., Aulehla, A., Gan, L. & Johnson, R. L. Lunatic fringe is an essential mediator of somite segmentation and patterning. *Nature* **394**, 377–381 (1998).
- Zhang, N. & Gridley, T. Defects in somite formation in lunatic fringe-deficient mice. *Nature* **394**, 374–377 (1998).
- Dale, J. K. et al. Periodic notch inhibition by lunatic fringe underlies the chick segmentation clock. *Nature* **421**, 275–278 (2003).
- Geling, A., Steiner, H., Willem, M., Bally-Cuif, L. & Haass, C. A γ -secretase inhibitor blocks Notch signaling *in vivo* and causes a severe neurogenic phenotype in zebrafish. *EMBO Rep.* **3**, 688–694 (2002).

18. Gibson, G. Developmental evolution: getting robust about robustness. *Curr. Biol.* **12**, R347–R349 (2002).
19. Stern, C. D. *et al.* Activin and its receptors during gastrulation and the later phases of mesoderm development in the chick embryo. *Dev. Biol.* **172**, 192–205 (1995).
20. Levin, M., Johnson, R. L., Stern, C. D., Kuehn, M. & Tabin, C. A molecular pathway determining left–right asymmetry in chick embryogenesis. *Cell* **82**, 803–814 (1995).
21. Boettger, T., Wittler, L. & Kessel, M. FGF8 functions in the specification of the right body side of the chick. *Curr. Biol.* **9**, 277–280 (1999).
22. Hirokawa, N. Stirring up development with the heterotrimeric kinesin KIF3. *Traffic* **1**, 29–34 (2000).
23. Wagner, M. K. & Yost, H. J. Left–right development: the roles of nodal cilia. *Curr. Biol.* **10**, R149–R151 (2000).
24. Levin, M., Thorlin, T., Robinson, K., Nogi, T. & Mercola, M. Asymmetries in H⁺/K⁺-ATPase and cell membrane potentials comprise a very early step in left–right patterning. *Cell* **111**, 77–89 (2002).
25. Stern, C. D. & Wolpert, L. Left–right asymmetry: all hands to the pump. *Curr. Biol.* **12**, R802–R803 (2002).
26. Mercola, M. Left–right asymmetry: nodal points. *J. Cell Sci.* **116**, 3251–3257 (2003).
27. Robinson, K. R. & Messerli, M. A. Left/right, up/down: The role of endogenous electrical fields as directional signals in development, repair and invasion. *Bioessays* **25**, 759–766 (2003).
28. Rao, Z. *et al.* The structure of a Ca²⁺-binding epidermal growth factor-like domain: its role in protein–protein interactions. *Cell* **82**, 131–141 (1995).
29. Rand, M. D., Lindblom, A., Carlson, J., Villoutreix, B. O. & Stenflo, J. Calcium binding to tandem repeats of EGF-like modules. Expression and characterization of the EGF-like modules of human Notch-1 implicated in receptor–ligand interactions. *Protein Sci.* **6**, 2059–2071 (1997).
30. McGrath, J., Somlo, S., Makova, S., Tian, X. & Brueckner, M. Two populations of node monocilia initiate left–right asymmetry in the mouse. *Cell* **114**, 61–73 (2003).
31. Hicks, C. *et al.* Fringe differentially modulates Jagged1 and Delta1 signalling through Notch1 and Notch2. *Nature Cell Biol.* **2**, 515–520 (2000).
32. Jarriault, S. *et al.* Signalling downstream of activated mammalian Notch. *Nature* **377**, 355–358 (1995).
33. Rand, M. D. *et al.* Calcium depletion dissociates and activates heterodimeric notch receptors. *Mol. Cell Biol.* **20**, 1825–1835 (2000).
34. Jaffe, L. F. Organization of early development by calcium patterns. *Bioessays* **21**, 657–667 (1999).
35. New, D. A. T. A new technique for the cultivation of the chick embryo *in vitro*. *J. Embryol. Exp. Morphol.* **3**, 326–331 (1955).
36. Ryan, A. K. *et al.* Ptx2 determines left–right asymmetry of internal organs in vertebrates. *Nature* **394**, 545–551 (1998).
37. Itasaki, N., Bel-Vialar, S. & Krumlauf, R. ‘Shocking’ developments in chick embryology: electroporation and *in ovo* gene expression. *Nature Cell Biol.* **1**, E203–E207 (1999).
38. Izpisua Belmonte, J. C., De Robertis, E. M., Storey, K. G. & Stern, C. D. The homeobox gene goosecoid and the origin of organizer cells in the early chick blastoderm. *Cell* **74**, 645–659 (1993).
39. Rodríguez-Esteban, C. *et al.* The novel Cer-like protein Caronte mediates the establishment of embryonic left–right asymmetry. *Nature* **401**, 243–251 (1999).
40. Laufer, E., Nelson, C. E., Johnson, R. L., Morgan, B. A. & Tabin, C. Sonic hedgehog and Fgf-4 act through a signaling cascade and feedback loop to integrate growth and patterning of the developing limb bud. *Cell* **79**, 993–1003 (1994).
41. Talora, C., Sgroi, D. C., Crum, C. P. & Dotto, G. P. Specific down-modulation of Notch1 signaling in cervical cancer cells is required for sustained HPV-E6/E7 expression and late steps of malignant transformation. *Genes Dev.* **16**, 2252–2263 (2002).
42. Lindsell, C. E., Shawber, C. J., Boulter, J. & Weinmaster, G. Jagged: a mammalian ligand that activates Notch1. *Cell* **80**, 909–917 (1995).
43. Shi, S. & Stanley, P. Protein O-fucosyltransferase 1 is an essential component of Notch signaling pathways. *Proc. Natl Acad. Sci. USA* **100**, 5234–5239 (2003).
44. Panin, V. M., Papayannopoulos, V., Wilson, R. & Irvine, K. D. Fringe modulates Notch–ligand interactions. *Nature* **387**, 908–912 (1997).
45. Bruckner, K., Perez, L., Clausen, H. & Cohen, S. Glycosyltransferase activity of Fringe modulates Notch-Delta interactions. *Nature* **406**, 411–415 (2000).
46. Moloney, D. J. *et al.* Fringe is a glycosyltransferase that modifies Notch. *Nature* **406**, 369–375 (2000).
47. Murray, J. D. *Mathematical Biology* (Springer, Berlin, 1993).
48. Rosenberg, M. W. & Massague, J. Juxtacrine cell signaling molecules. *Curr. Opin. Cell Biol.* **5**, 832–838 (1993).
49. Collier, J. R., Monk, N. A., Maini, P. K. & Lewis, J. H. Pattern formation by lateral inhibition with feedback: a mathematical model of delta-notch intercellular signalling. *J. Theor. Biol.* **183**, 429–446 (1996).
50. Owen, M. R., Sherratt, J. A. & Wearing, H. J. Lateral induction by juxtacrine signaling is a new mechanism for pattern formation. *Dev. Biol.* **217**, 54–61 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M-F. Schwarz and H. Pineda for their technical skills and dedication; A. Lehrman and H. Juguilon for help with the luciferase assays; G. Sternik for assistance in two-photon excitation microscopy analysis; all laboratory members for discussions; L. Hooks for help in preparing the manuscript; C. Fryers, K. Jones and C. Kintner for reagents and discussions; and T. Gridley and R. Johnson for sharing unpublished results. A.R. is partially supported by a postdoctoral fellowship from the Ministerio de Educación, Cultura y Deporte, Spain; J.R.L. is supported by a fellowship from Fundação para a Ciência e a Tecnologia, Portugal, and M.I. is partially supported by the Fulbright Program and Generalitat of Catalunya. This work was supported by grants from the Fundação Calouste Gulbenkian e Fundação para a Ciência e a Tecnologia, the American Heart Association, the Human Frontier Science Program, the NIH and the G. Harold and Leila Y. Mathers Charitable Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.C.I.B. (belmonte@salk.edu).

The massive binary companion star to the progenitor of supernova 1993J

Justyn R. Maund¹, Stephen J. Smartt¹, Rolf P. Kudritzki², Philipp Podsiadlowski³ & Gerard F. Gilmore¹

¹*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge CB3 0HA, UK*

²*Institute for Astronomy, University of Hawaii at Manoa, 2860 Woodlawn Drive, Honolulu, Hawaii 96822, USA*

³*University of Oxford, Astrophysics, Keble Road, Oxford OX1 3RH, UK*

The massive star that underwent a collapse of its core to produce supernova (SN)1993J was subsequently identified as a non-variable red supergiant star in images of the galaxy M81 taken before explosion^{1,2}. It showed an excess in ultraviolet and B-band colours, suggesting either the presence of a hot, massive companion star or that it was embedded in an unresolved young stellar association¹. The spectra of SN1993J underwent a remarkable transformation from the signature of a hydrogen-rich type II supernova to one of a helium-rich (hydrogen-deficient) type Ib^{3,4}. The spectral and photometric peculiarities were best explained by models in which the 13–20 solar mass supergiant had lost almost its entire hydrogen envelope to a close binary companion^{5–7}, producing a ‘type IIB’ supernova, but the hypothetical massive companion stars for this class of supernovae have so far eluded discovery. Here we report photometric and spectroscopic observations of SN1993J ten years after the explosion. At the position of the fading supernova we detect the unambiguous signature of a massive star: the binary companion to the progenitor.

Images of SN1993J were acquired on 28 May 2002 using the Advanced Camera for Surveys (ACS) on board the Hubble Space Telescope (HST). These observations were primarily carried out in the ultraviolet and blue spectral regions, where the flux of a hot star companion would be most prominent. A F330W (U-band) image of SN1993J is shown in Fig. 1. The high angular resolution of these images (0.025'' per pixel) allowed for the complete separation of unrelated surrounding stars from the supernova, including the discovery of a previously blended star near the supernova (star G of Fig. 1; see Table 1 for details). The blending of stars E–G with the supernova (as in previous lower-resolution HST images⁸) would lead to the supernova appearing bluer. The stars surrounding the supernova, which are isolated by the exquisite resolution of the ACS images, would be blended with the supernova flux in lower-spatial-resolution ground-based observations. Recently the contributions of the surrounding blue stars to the U- and B-band magnitudes of the supernova progenitor have been estimated from lower-resolution HST archive images⁸. This study concluded that although a significant fraction of the progenitor U-band flux could be explained by the contamination of these stars in the ground-based point spread function, within 1 σ errors, a hot massive companion star would be compatible with the progenitor and the late-time images. We have conducted a similar analysis with the higher-spatial-resolution and deeper ACS images. A gaussian weighting scheme, which is a function of angular distance of the stars from the supernova and the wavelength-dependent seeing, was applied to simulate the contribution of each object to the photometry of the supernova progenitor¹. We found that these stars, assuming a K0Ia progenitor, clearly would not have had sufficient flux to be solely responsible for the ultraviolet excess reported for the progenitor^{1,8}. In a similar fashion this photometry provided an estimate of the flux contribution of these surrounding stars to the ground-based spectroscopy of SN1993J.

Spectroscopy of SN1993J was acquired using the LRIS-B

spectrometer on the Keck I telescope, covering 3,400–5,400 Å, on 1 March 2003 (Fig. 2). The spectrum in general is typical of supernovae at late times that show evidence for the interaction of the ejecta with a dense circumstellar medium⁹. It is dominated by a series of broad, box-like emission lines, which have been interpreted to be due to ejecta shocks in a thin spherical shell¹⁰. However, the high signal-to-noise ratio of the LRIS-B spectrum in the near-ultraviolet illustrates two peculiar features not detected before: a rising ultraviolet continuum and narrow absorption lines consistent with a hot star spectrum superimposed on the flux from the young supernova remnant. The maximum contribution of the surrounding stars to the observed spectrum, again simulating the effects of ground-based seeing conditions with a gaussian weighting function, was calculated to be 18%. The variability of star A^{2,8} (see Fig. 1), given its magnitude and distance from the supernova, has a negligible effect on this analysis. The strength and width of the absorption lines, and the rising shape of the blue continuum suggest that the observed flux is composed of a pseudo-continuum from the supernova ejecta emission and flux from a hot star that is unresolved from the supernova on the ACS images. The point spread function of SN1993J is perfectly consistent with a single stellar source, and conservatively we have estimated that the separation of the star and the supernova must be closer than 0.015 arcseconds (0.3 parsecs), or else we would clearly detect asymmetries in the spatial profile. The observed B-band stellar density of M81 in the vicinity of SN1993J implied a probability of 1 in 2,000 of a single randomly located and unrelated star fortuitously lying so close to the supernova as to be unresolved.

The narrow stellar lines observed in the SN1993J spectrum were compared with lines from observed spectra of late-O and early-B standard stars and TLUSTY^{11,12} model stellar spectra, with appropriate parameters^{13,14}. The comparison was conducted in the wavelength region of the U-band, least affected by the strong supernova

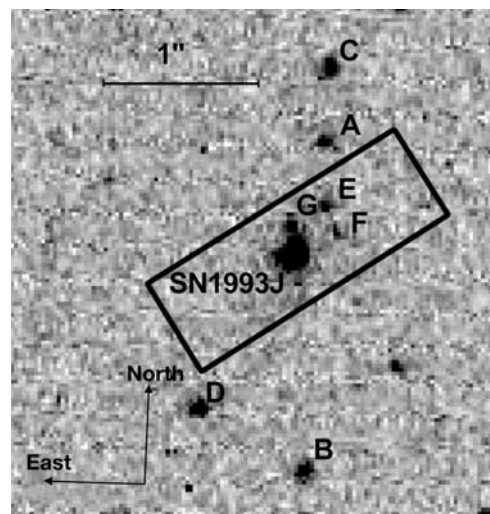


Figure 1 An HST ACS High Resolution Camera (HRC) F330W (~U-band) image of SN1993J. The image was a 1,200-s exposure, acquired on 28 May 2002—9.17 yr post-explosion, observed for program G09353. The spatial resolution of the HRC allows the clear separation of stars E–G from the supernova (adopting the nomenclature of ref. 8). The photometric magnitudes of the stars measured in the ACS filters²⁴ are listed in Table 1. These magnitudes were used to calculate, and remove, the contamination of the faint surrounding blue stars to the ground-based spectrum of SN1993J. The box is an example of the area of sky sampled in the lower-spatial-resolution ground-based spectroscopy. The width is 0.7'' (the slit width employed with the LRIS spectrograph), and the length is 1.8'' (the aperture diameter used to extract the object spectrum down to a level of 10% of the peak flux). The slit was positioned at the angle of parallax and during the night rotated between angles -162° and 106° (east of north). The optical spectrum presented in Fig. 2 is effectively a sum of the flux of all objects within this box.

Table 1 Photometric magnitudes of SN1993J and nearby stars

Star	Distance from supernova (pc)	Near-ultraviolet (F250W)		U		B		V	
A	12.9	22.99	(0.05)	23.05	(0.08)	23.78	(0.02)	22.56	(0.01)
B	25.0	22.76	(0.05)	22.89	(0.06)	23.12	(0.01)	23.05	(0.01)
C	21.1	22.52	(0.05)	22.98	(0.08)	23.78	(0.01)	23.80	(0.01)
D	20.8	22.15	(0.06)	22.87	(0.07)	23.84	(0.01)	23.95	(0.01)
E	6.0	23.39	(0.07)	23.68	(0.08)	24.12	(0.01)	24.17	(0.01)
F	5.1	23.73	(0.12)	24.2	(0.18)	25.04	(0.01)	24.85	(0.01)
G	2.8	22.79	(0.19)	23.54	(0.08)	24.44	(0.02)	24.42	(0.01)
1993J	—	19.93	(0.04)	20.7	(0.05)	21.08	(0.01)	20.27	(0.01)

Star numbers are as identified on Fig. 1, using the nomenclature of ref. 8. The distances of the stars, in parsec, are calculated from the centre of the SN1993J point spread function. Magnitudes are F250W Vegamag, in the ACS photometric system²⁴ and U, B and V Johnson magnitudes, transformed from the original ACS photometric system. Bracketed numbers indicate the uncertainty, in magnitudes, of the preceding magnitude. The colours of the surrounding stars imply they are of early type, in the range O9–B3. Photometry of 24 other stars, within 106 pc of SN1993J, were used to determine the reddening by comparing observed colours (U–B) and (B–V) with intrinsic colours, yielding a colour excess $E(B-V) = 0.2$ and hence a V-band extinction $A_V = 0.62$, assuming a standard galactic reddening law, which is in good agreement with previously determined values¹⁰. An accurate distance to the galaxy of 3.63 ± 0.14 Mpc from Cepheid variables is available²⁵.

emission lines. Main-sequence stars were immediately excluded because their luminosities were too low to produce lines of sufficient strength to match those in the SN1993J spectrum. The strengths of the lines in the SN1993J spectrum, with respect to the continuum, were weaker than those in the stellar spectra owing to the continuum of the supernova. The continuum of the supernova, in units of the B supergiant flux at each H I line, was calculated by adding excess continuum to the stellar spectra until the renormalized H I line strengths matched those observed in the SN1993J spectrum. In addition, the continua ratio was calculated for He I lines, in order to

identify those types which gave consistent matches with the values calculated for the H I lines assuming a normal He abundance. B-type supergiants of type B5 and later were eliminated by this method. Early B-supergiants (B0–B4) provided the best consistent fit to the H I and He I lines. The average ratio of the continua, for all the H I lines, was calculated for each spectral type: for B0Ia it was 1, for B1Ia 1.9, for B2Ia 1.9, for B3Ia 2.3 and for B4Ia 3 (where the contribution due to the surrounding stars is already subtracted). From the ACS U- and B-band fluxes of the supernova plus the companion, and the relative fluxes of the companion required to

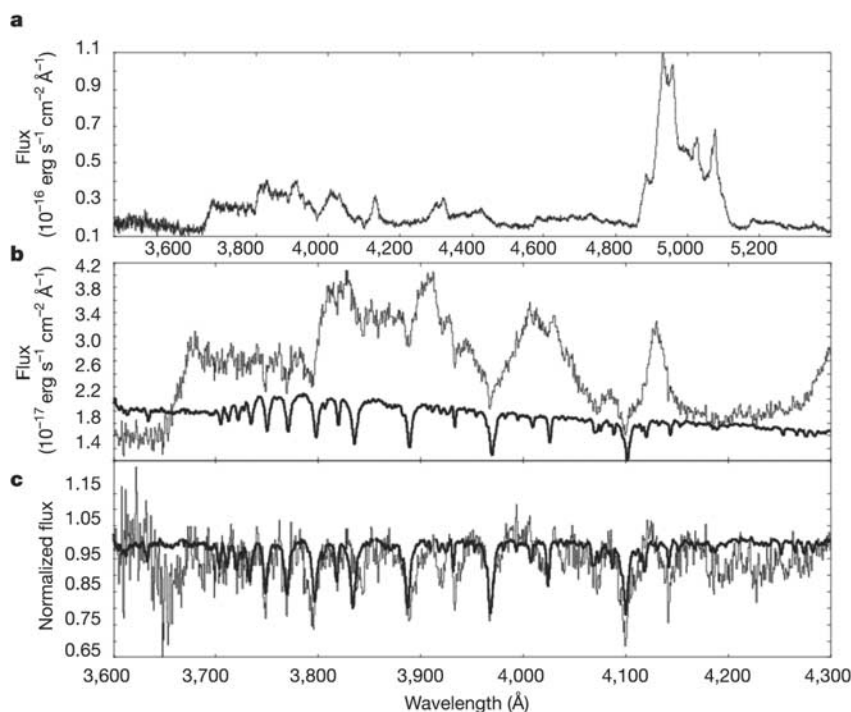


Figure 2 Near-ultraviolet spectra of SN1993J. **a**, Flux-calibrated ground-based blue spectrum of SN1993J. The spectrum was acquired with Keck I LRIS-B spectrometer (1 March 2003, 9.93 yr post-explosion) using the 600/4,000 grism (with instrumental resolution 2.4 Å and signal-to-noise ratio ranging from 15–30). Seeing during the 5.5-hr total exposure time was between 0.7 and 1.2 arcseconds. The spectrum is dominated by broad, box-like emission line profiles which are characteristic of interaction between the supernova ejecta and the circumstellar medium^{9,10}. However, the near-ultraviolet region shows a series of narrow absorption features. **b**, Flux spectra of SN1993J (thin line) and a B-supergiant (bold line) HD168489 (B1Ia, $T_{\text{eff}} \approx 23.3 \times 10^3 \text{ K}$ and $\log L/L_{\odot} = 5$). The stellar spectrum was scaled to the distance of M81, with a reddening of $E(B-V) = 0.2$ and a radial velocity shift of -120 km s^{-1} . The series of sharp absorption features are coincident with the H I Balmer series and the He I line at $3,819.7 \text{ Å}$. The absorption lines are conspicuously narrower than the [O III] (wavelengths, $\lambda = 4,959$ and $5,007 \text{ Å}$) and

[S II] ($\lambda = 4,068$ and $4,072 \text{ Å}$) emission lines ($3,200 \text{ km s}^{-1}$)—implying that the absorption lines do not arise from the supernova ejecta itself. The flux contribution from the contaminating stars A–G (18%) cannot reproduce the strength of the absorption lines observed. **c**, Normalized spectra of SN1993J (thin line) and the B1Ia Galactic standard star HD168489, which has been renormalized for an excess continuum ratio of 0.8, in units of the combined continuum fluxes of the companion and surrounding stars. The supernova continuum was normalized by estimating the position of the continuum and applying a series of short spline fits. The H I Balmer and He I absorption lines are best matched by early-B type supergiant spectra (B0–B4). The average ratios, in the range $3,600\text{--}4,000 \text{ Å}$, and the total measured ACS flux in the F330W filter allowed a calculation¹⁴ of the parameters of the unresolved hot star component: $\log L/L_{\odot} = 5 \pm 0.3$, $\log T_{\text{eff}} = 4.3 \pm 0.1$.

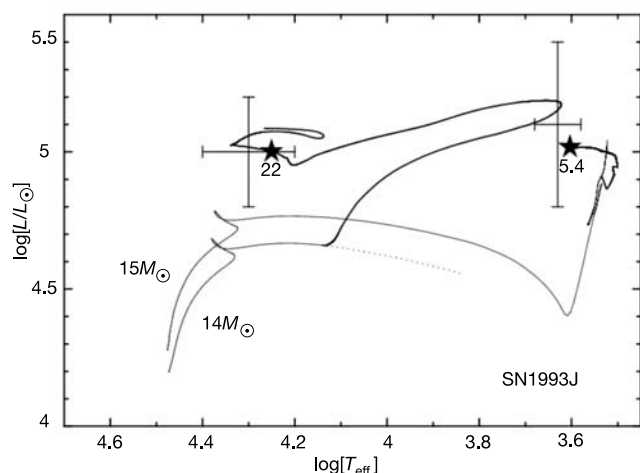


Figure 3 Hertzsprung–Russell diagram illustrating the evolution of the progenitor binary of SN1993J. The system initially consists of two stars of mass $15M_{\odot}$ and $14M_{\odot}$ in an orbit of 5.8 yr. The thin lines show the evolution of both components before the beginning of mass transfer, the bold lines during the mass-transfer phase. The more massive component fills its Roche lobe and starts to transfer mass to the companion after the exhaustion of central helium when ascending the red-supergiant branch for the second time (called ‘case C’ mass transfer; alternatively mass transfer may already start on the first ascent of the red-supergiant branch; ‘case BC’ mass transfer). The mass transfer rate is initially very high, reaching a peak of $4 \times 10^{-2} M_{\odot} \text{ yr}^{-1}$ (the accretion rate onto the secondary was limited to $2 \times 10^{-2} M_{\odot} \text{ yr}^{-1}$). Once the primary has transferred most of its envelope, it detaches from its Roche lobe and subsequently continues to lose mass owing to a strong stellar wind, assumed to be $4 \times 10^{-5} M_{\odot} \text{ yr}^{-1}$ (ref. 25) at the time of the explosion. Because of the similar masses, the secondary is already somewhat evolved and is near the end of its hydrogen-core burning phase or just beyond (and hence is not being rejuvenated by accretion^{15,19}). The numbers along the tracks give the masses of the two components at selected phases. At the time of the explosion of the primary (indicated by asterisks), the primary has a mass of $5.4M_{\odot}$ (with a helium-exhausted core of $5.1M_{\odot}$), the secondary has a mass of $22M_{\odot}$. The error bars show the parameters of the two components, as determined in this work.

produce these narrow lines, the luminosity and temperature of the companion star were calculated to be: $\log L/L_{\odot} = 5 \pm 0.3$ and $\log T_{\text{eff}} = 4.3 \pm 0.1$ (where subscript $\odot$ refers to the Sun, and our best estimate is a B2Ia star). The uncertainties were determined from the range of atmospheric parameters that would provide appropriate simultaneous matches to the observed lines. The pre-explosion photometry was used to further constrain the companion parameters. The luminosity of the B0Ia star requires a U-band magnitude too bright to be compatible with the pre-explosion U-band luminosity of the progenitor binary system. The nature of the progenitor, considering the blend of the progenitor and the companion in the pre-explosion photometry, is restricted to be a K-supergiant with parameters $\log L/L_{\odot} = 5.1 \pm 0.3$ and $\log T_{\text{eff}} = 3.63 \pm 0.05$.

The most important implication of these results is that it confirms that the progenitor of SN1993J was a member of a binary system which has lost most of its hydrogen-rich envelope by interaction with its companion star^{5–7}. Figure 3 illustrates the evolutionary tracks of both components, the mass-losing one and the mass-accreting one, in the Hertzsprung–Russell (H–R) diagram. The model is very similar to the ones proposed originally^{5–7}. The system initially consists of two stars of comparable masses (15 and $14M_{\odot}$) with an orbital period of 5.8 yr. At the time of the supernova explosion the locations of both stars in the H–R diagram are close to the observationally inferred ones. Note that the subsequent evolution of the companion star would be very different from a star evolved in isolation, and it would most probably end its evolution as

a blue supergiant rather than a red supergiant, producing a supernova like SN1987A^{15–18} (see Fig. 3). An overabundance of He would be a signature of the accretion onto the companion from the progenitor and the supernova, and we would expect that the He abundance could be determined at much later times once the supernova has faded sufficiently^{19,20}.

The narrow lines in the SN1993J spectrum are at a velocity of $\sim 120 \text{ km s}^{-1}$ relative to the rest frame. This velocity is consistent with the rotational velocity curve of M81 at the position of SN1993J^{10,21} and with a low kick velocity to the companion in a long-period binary. At the time of the explosion the orbital period of the system is ~ 25 yr, and the companion has an orbital velocity of $\sim 6 \text{ km s}^{-1}$. The companion is expected to move with this velocity relative to its neighbourhood if the system is disrupted as a result of the supernova, or with the post-supernova binary system velocity ($\sim 20\text{--}40 \text{ km s}^{-1}$) if the system remains bound²². □

Received 29 August; accepted 23 October 2003; doi:10.1038/nature02161.

1. Aldering, G., Humphreys, R. M. & Richmond, M. SN1993J: The optical properties of its progenitor. *Astron. J.* **107**, 662–672 (1994).
2. Cohen, J. G., Darling, J. & Porter, A. The nonvariability of the progenitor of supernova 1993J in M81. *Astron. J.* **110**, 308–311 (1995).
3. Filippenko, A. V., Matheson, T. & Ho, L. C. The ‘‘type IIb’’ supernova 1993J in M81: a close relative of type Ib supernovae. *Astrophys. J.* **415**, L103–L106 (1993).
4. Swartz, D. A., Clacchiatti, A., Benjamin, R., Lester, D. F. & Wheeler, J. C. Supernova 1993J as a spectroscopic link between type II and type Ib supernovae. *Nature* **365**, 232–234 (1993).
5. Podsiadlowski, Ph., Hsu, J. J. L., Joss, P. C. & Ross, R. R. The progenitor of supernova 1993J: a stripped supergiant in a binary system. *Nature* **364**, 509–511 (1993).
6. Nomoto, K. et al. A type IIb model for supernova 1993J. *Nature* **364**, 507–509 (1993).
7. Woosley, S. E., Eastman, R. G., Weaver, T. A. & Pinto, P. A. SN1993J: A type IIb supernova. *Astrophys. J.* **429**, 300–318 (1994).
8. Van Dyk, S. D. et al. The progenitor of supernova 1993J revisited. *Publ. Astron. Soc. Pacif.* **114**, 1322–1332 (2002).
9. Fesen, R. A. et al. Late-time optical and ultraviolet spectra of SN1979C and SN1980K. *Astron. J.* **117**, 725–735 (1999).
10. Matheson, T., Filippenko, A. V., Ho, L. C., Barth, A. J. & Leonard, D. C. Detailed analysis of early to late-time spectra of supernova 1993J. *Astron. J.* **120**, 1499–1515 (2000).
11. Hubeny, I. & Lanz, T. Non-LTE line-blanketed model atmospheres of hot stars. 1: Hybrid complete linearization/accelerated lambda iteration method. *Astrophys. J.* **439**, 875–904 (1995).
12. Ryans, R. et al. An analysis of the optical spectra of the post-asymptotic giant branch stars LSIV-12 111 and HD 341617. *Astron. Astrophys.* **401**, 1119–1129 (2003).
13. McErlean, N. D., Lennon, D. J. & Dufton, P. L. Galactic B-supergiants: A non-LTE model atmosphere analysis to estimate atmospheric parameters and chemical compositions. *Astron. Astrophys.* **349**, 553–572 (1999).
14. Kudritzki, R. et al. The wind momentum–luminosity relationship of galactic A- and B-supergiants. *Astron. Astrophys.* **350**, 970–984 (1999).
15. Podsiadlowski, Ph. & Joss, P. C. An alternative binary model for SN1987A. *Nature* **338**, 401–403 (1989).
16. Braun, H. & Langer, N. Effects of accretion onto massive main sequence stars. *Astron. Astrophys.* **297**, 483–493 (1995).
17. Podsiadlowski, Ph. The progenitor of SN 1987A. *Publ. Astron. Soc. Pacif.* **104**, 717–729 (1992).
18. Smartt, S. J. et al. The evolutionary status of SN 25—implications for blue supergiants and the progenitor of SN1987A. *Astron. Astrophys.* **391**, 979–991 (2002).
19. Bresolin, F. et al. Discovery and quantitative spectral analysis of an Ofpe/WN9 (WN11) star in the sculptor spiral galaxy NGC 300. *Astrophys. J.* **577**, L107–L110 (2002).
20. Urbaneja, M. et al. Quantitative spectral analysis of early B-type supergiants in the sculptor galaxy NGC 300. *Astrophys. J.* **584**, L73–L77 (2003).
21. Marggraf, O. & de Boer, K. S. Absorption line systems on the line of sight from SN 1987A and SN1993J and the intergalactic cloud in front of M81. *Astron. Astrophys.* **363**, 733–743 (2000).
22. Brandt, N. & Podsiadlowski, Ph. The effects of high-velocity supernova kicks on the orbital properties and sky distributions of neutron star binaries. *Mon. Not. R. Astron. Soc.* **274**, 461–484 (1995).
23. Ferrarese, L. et al. A database of Cepheid distance moduli and tip of the red giant branch, globular cluster luminosity function, planetary nebula luminosity function and surface brightness fluctuations data useful for distance determinations. *Astrophys. J. Suppl. Ser.* **128**, 431–459 (2000).
24. Pavlovsky, C. et al. *ACS Instrument Handbook Version 3.0 99* (STScI, Baltimore, 2002).
25. Fransson, C., Lundqvist, P. & Chevalier, R. A. Circumstellar interaction in SN1993J. *Astrophys. J.* **461**, 993–1008 (1996).

Acknowledgements This paper is based on observations made with the NASA/ESA Hubble Space Telescope and the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California, the National Aeronautics and Space Administration and the University of Hawaii. We thank R. Ryans for providing us with new TLUSTY calculations of the hydrogen line spectra for B-type supergiants. J.R.M. and S.J.S. thank PPARC for financial support and thank A. D. Mackey and J. Mack for advice concerning photometric data analysis.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R.M. (jrm@ast.cam.ac.uk).

An intense stratospheric jet on Jupiter

F. M. Flasar¹, V. G. Kunde², R. K. Achterberg³, B. J. Conrath⁴, A. A. Simon-Miller¹, C. A. Nixon², P. J. Gierasch⁴, P. N. Romani¹, B. Bézard⁵, P. Irwin⁶, G. L. Bjoraker¹, J. C. Brasunas¹, D. E. Jennings¹, J. C. Pearl¹, M. D. Smith¹, G. S. Orton⁷, L. J. Spilker⁷, R. Carlson³, S. B. Calcutt⁶, P. L. Read⁶, F. W. Taylor⁶, P. Parrish⁶, A. Barucci⁵, R. Courtin⁵, A. Coustenis⁵, D. Gautier⁵, E. Lellouch⁵, A. Marten⁵, R. Prangé⁵, Y. Biraud⁵, T. Fouchet⁵, C. Ferrari⁸, T. C. Owen⁹, M. M. Abbas¹⁰, R. E. Samuelson², F. Raulin¹¹, P. Ade¹², C. J. Césarsky¹³, K. U. Grossman¹⁴ & A. Coradini¹⁵

¹NASA/Goddard Space Flight Center, Code 693, Greenbelt, Maryland 20771, USA

²Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA

³Science Systems and Applications, Inc., 5900 Princess Garden Parkway, Suite 300, Lanham, Maryland 20706, USA

⁴Department of Astronomy, Cornell University, Ithaca, New York 14853, USA

⁵LESIA, CNRS-FRE 2461, Observatoire de Paris, 5 place Jules Janssen,

F-91925 Meudon Cedex, France

⁶Atmospheric, Oceanic, and Planetary Physics, Clarendon Laboratory, Parks Rd, University of Oxford, Oxford OX1 3PU, UK

⁷Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, California 91109, USA

⁸CEA Saclay, Service d'Astrophysique, 91191 Gif-sur-Yvette Cedex, France

⁹University of Hawaii, Institute of Astronomy, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

¹⁰NASA/Marshall Space Flight Center, SD50 NSSTC, Huntsville, Alabama 35812, USA

¹¹LISA, Université de Paris 7 & 12, CNRS – UMR 7583, 61 Ave. General de Gaulle, 94010 Creteil Cedex, France

¹²Department of Physics and Astronomy, University of Cardiff, 5 The Parade, Cardiff CF24 3YB, UK

¹³European Southern Observatory, Karl-Schwarzschild-strasse 2, 85748 Garching bei Muenchen, Germany

¹⁴Gesamthochschule Wuppertal, Department of Physics, Gausstrasse 20, 5600 Wuppertal 1, Germany

¹⁵Istituto di Astrofisica Spaziale – CNR, Area della ricerca di Tor Vergata, Via del Fosso del Cavaliere 100, Rome, I-00133 Italy

The Earth's equatorial stratosphere shows oscillations in which the east–west winds reverse direction and the temperatures change cyclically with a period of about two years^{1,2}. This phenomenon, called the quasi-biennial oscillation, also affects the dynamics of the mid- and high-latitude stratosphere and weather in the lower atmosphere². Ground-based observations have suggested^{3–5} that similar temperature oscillations (with a 4–5-yr cycle) occur on Jupiter, but these data suffer from poor vertical resolution and Jupiter's stratospheric wind velocities have not yet been determined. Here we report maps of temperatures and winds with high spatial resolution, obtained from spacecraft measurements of infrared spectra of Jupiter's stratosphere. We find an intense, high-altitude equatorial jet with a speed of $\sim 140 \text{ m s}^{-1}$, whose spatial structure resembles that of a quasi-quadrennial oscillation. Wave activity in the stratosphere also appears analogous to that occurring on Earth. A strong interaction between Jupiter and its plasma environment produces hot spots in its upper atmosphere and stratosphere near its poles^{6–9}, and the temperature maps define the penetration of the hot spots into the stratosphere.

Jupiter's banded system, with alternating eastward and westward winds, has been extensively mapped at the visible cloud level from time-lapsed images^{10–13}. Few distinct tracers are observed at higher altitudes, and little has been known about the winds there, particularly in the stratosphere, which begins approximately 25 km above

the visible clouds. Jupiter rotates rapidly, like Earth, and the thermal wind relation¹⁴ can be used to determine the winds above the clouds, provided that the temperatures there are known with adequate spatial resolution. However, most of the previous determinations of stratospheric temperatures have come from ground-based observations^{3,5} that have had poor vertical resolution, because

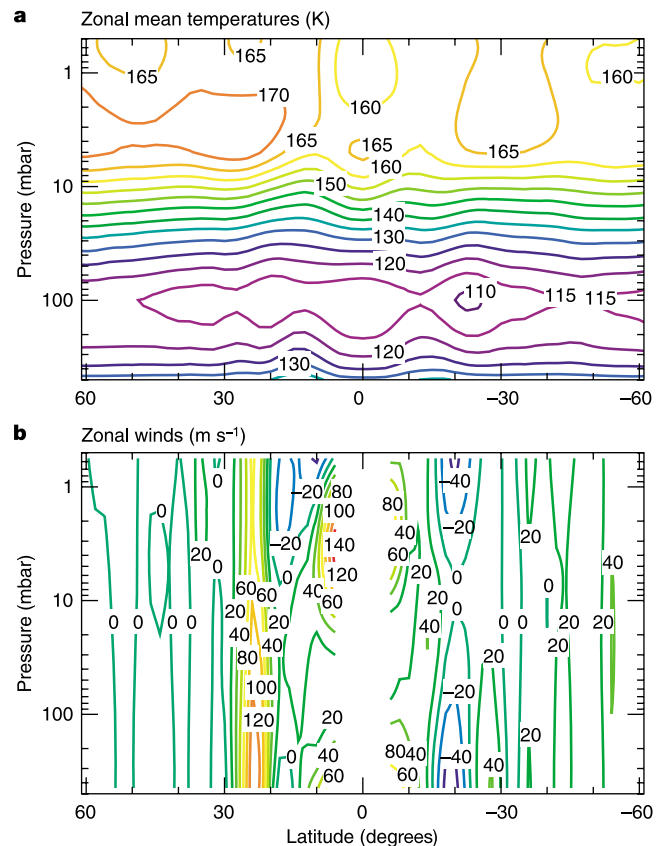


Figure 1 Temperatures and winds above Jupiter's clouds show the high-altitude stratospheric jet at the equator. **a**, Atmospheric temperatures, and **b**, zonal (east–west) winds, averaged over longitude, as functions of pressure as the vertical coordinate and latitude as the horizontal coordinate. Latitude is planetographic^{25,26}. Mid-infrared spectra (600–1,400 cm^{-1}) were used to construct the cross-sections. The spectra used were obtained just after Jupiter closest approach, and the spatial resolution was 2.4° of latitude at the subspacecraft point, near the equator and degraded as $1/\cos(\text{latitude})$ elsewhere. Vertical profiles of atmospheric temperature were retrieved by means of a constrained inversion with strong low-pass filtering²⁷. The known opacities of the collision-induced S(1) line of H_2 near 600 cm^{-1} and the vibrational-rotational band of CH_4 near $1,300 \text{ cm}^{-1}$ permitted the retrieval of temperatures between 80 and 400 mbar in the troposphere and between 1 and 20 mbar in the stratosphere, respectively. Stratospheric temperatures at intermediate pressures were interpolated. The resolution of CIRS spectra can be commanded between 0.5 and 16 cm^{-1} , and the temperatures presented are from 3-cm^{-1} spectra. The zonal winds have been derived from the temperatures displayed, using the thermal wind equation, which relates the vertical gradient of the zonal winds to the north–south temperature gradient in rapidly rotating fluids¹⁴. Zonal wind velocities, determined from tracking distinct features at the visible cloud level from time-lapsed images¹², have been used as a lower boundary condition at 500 mbar. (They have been smoothed to 3° latitude resolution, which tends to reduce the speeds from those actually measured.) Positive numbers denote eastward velocities. Equatorial latitudes are omitted because the thermal wind equation greatly amplifies the errors in temperature when the latitude is small. The error in temperature from instrument noise is 0.15 K, and this propagates to an error of $0.5/\sin(\text{latitude}) \text{ m s}^{-1}$ per scale height in the zonal winds. Contours are coloured for clarity. For temperatures in **a**, orange to purple span 170 K to 110 K; for winds in **b**, red to blue span largest eastward winds (140 m s^{-1}) to largest westward winds (-40 m s^{-1}).

of the limitations imposed by the Earth's atmospheric opacity. In effect, one obtains temperatures averaged over a single thick layer centred at 10–20 mbar.

We acquired spectra during the swing-by of the Cassini spacecraft past Jupiter in late 2000 and early 2001, using the Composite Infrared Spectrometer (CIRS), a Fourier-transform spectrometer described elsewhere¹⁵. These spectra allowed us to construct temperature maps with high vertical and horizontal resolution, and from these to determine the winds above the clouds. Figure 1 displays the latitude–height cross-sections of the zonal-mean temperature and zonal winds. The winds decay with height in the upper troposphere, in agreement with previous determinations¹⁶. Higher up, a more complex behaviour emerges. In some cases, such as the strong eastward jet at 23° N, the decay persists. In others, the winds build back up in strength. This occurs for the westward jet near 18° S, as well as for a few jets at higher latitudes, although with smaller amplitudes.

Of particular note are the temperature and wind fields at low latitudes. Near 100 mbar and 1 mbar, the equatorial temperatures are relatively colder than those at 10–15° N and S; at intermediate levels, 4–20 mbar, the equator is up to 5 K warmer. Concomitantly,

after an initial decay in the upper troposphere, the low-latitude zonal winds build up in the stratosphere, to form a $\sim 140 \text{ m s}^{-1}$ jet centred near 4 mbar. The presence of such an intense high-altitude jet was not previously known, and it rivals the magnitudes of the equatorial winds at the cloud tops, which are $\sim 100\text{--}150 \text{ m s}^{-1}$ (ref. 12). The stratospheric winds are at least as large as the figure indicates and may be larger, because the available spatial resolution may average larger meridional temperature gradients on smaller scales.

Comparison of our results with ground-based infrared observations suggests that the low-latitude temperature structure in Fig. 1 is a response to dynamical forcing. Despite their limited vertical resolution, these observations extend over two decades, and show that the low-latitude stratospheric temperatures oscillate with a 4–5-yr cycle^{3,5}. Leovy *et al.*⁴ have proposed that this behaviour, which they named the quasi-quadrennial oscillation, or QQO, is analogous to the Earth's quasi-biennial oscillation, in which stratospheric zonal winds at low latitudes alternate between eastward and westward with an approximate period of two years. The zonal-mean structure is modulated by the stresses of vertically propagating waves, and waves with eastward and westward phase

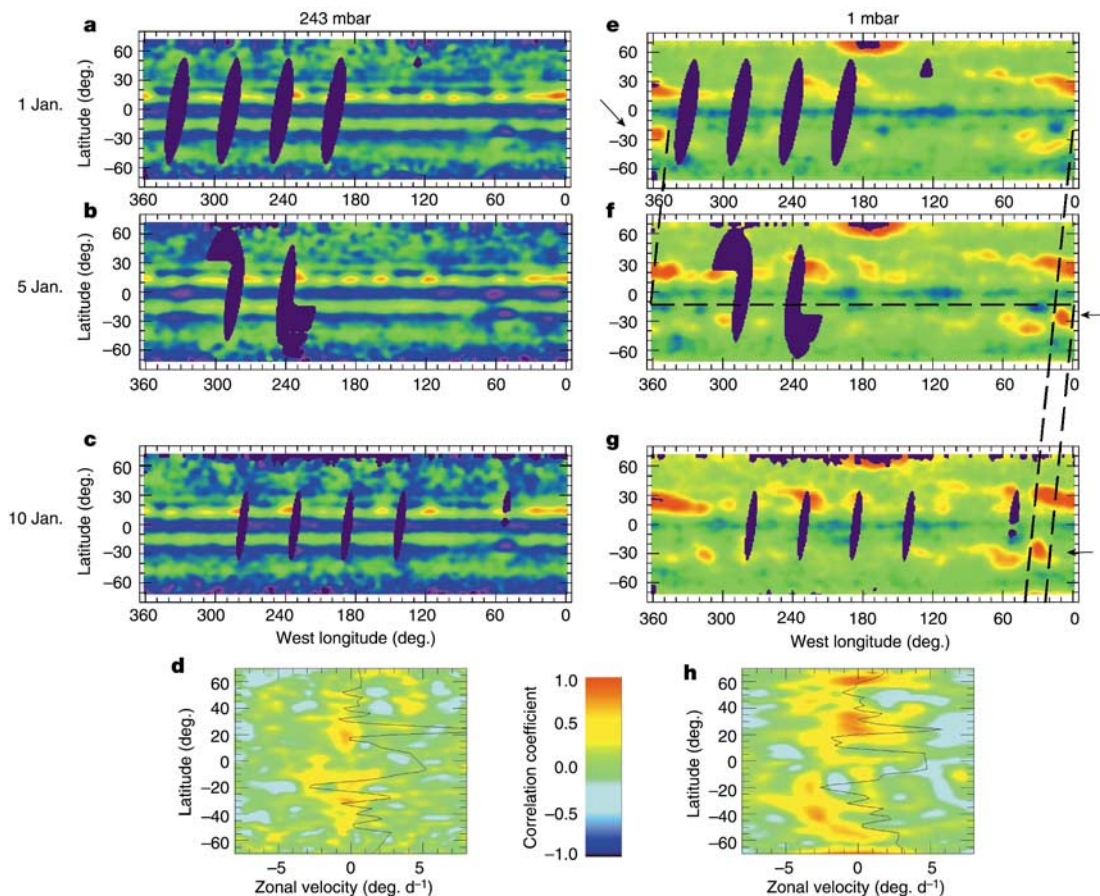


Figure 2 Maps illustrating the motions of thermal features at two levels in Jupiter's atmosphere. Temperatures and motions of thermal features in the upper troposphere (243 mbar; left panels) and upper stratosphere (1 mbar; right). **a–c, e–g**, Temperature maps from three global mapping sequences in January 2001. West longitudes are in System III^{25,28}, which should be indicative of Jupiter's interior rotation. The CIRS mid-infrared focal planes consist of two parallel 1×10 pixel arrays, each square pixel subtending 0.3 mrad on a side. The maps were obtained by repeatedly scanning the arrays in a push-broom fashion (perpendicular to the array lengths) from north to south along the subspacecraft meridian, as Jupiter rotated. The scanning rate was such that all longitudes would be mapped in two planetary rotations (~ 20 h). The spatial resolution of

each pixel near the equator ranged from 2.4° in latitude and longitude on 1 January to 3.5° on 10 January. Colour scale as follows: purple to red spans 106 K to 140 K in **a–c**, and 140 K to 180 K in **e–g**. Dark areas indicate gaps in the data coverage. The error in temperature from instrument noise, here and in Fig. 3, is ~ 1 K. The vertical spacing of the panels is proportional to the times of the maps. The arrows and dashed black lines highlight the 3.9° d^{-1} (50 m s^{-1}) westward drift of the stratospheric warm feature at 25° S, initially lying between 350° W and 360° W. **d, h**, The eastward zonal velocities implied by the cross-correlations of the thermal structure on 1 and 10 January. The solid curves superimposed are the mean zonal winds at the two pressure levels, taken from Fig. 1.

velocities are required to drive the oscillation². Despite the poor vertical resolution of the ground-based data, detailed modelling of the QJO based on the retrieved temperatures^{5,17} have predicted a spatial structure of the temperature and zonal winds that is qualitatively consistent with the CIRS observations. However, the vertical scale derived for the reversal in the zonal winds and the meridional contrasts of temperature are approximately twice as large as those indicated in Fig. 1. The models suggest that the oscillatory behaviour may penetrate to the upper troposphere, and the new data substantiate this. Figure 1 shows that equatorial temperatures at these altitudes are ~ 5 K colder than those at adjacent latitudes. This depression is significantly larger than that measured by the Voyager infrared spectrometer (IRIS) in 1979: ~ 2 K (ref. 16).

The nature of the waves forcing the equatorial oscillation on Jupiter is not known. The model predictions range from planetary-scale waves¹⁷ to gravity waves with horizontal wavelengths well below the CIRS spatial resolution⁵. In analogy to Earth¹, propagating waves probably play an important role in momentum transport and in modulating Jupiter's zonal winds, not only near the equator, but globally. To search for planetary-scale waves, we examine the zonal variation in temperatures. Figure 2a–c, e–g shows latitude–

longitude maps of atmospheric temperature in the troposphere and in the stratosphere. Within 15° of the equator there is zonal structure that may indicate wave activity associated with a QJO. However, the figure also indicates a variety of thermal features away from the equator. In the troposphere (Fig. 2a–c), long trains of regularly spaced features occur. From both visual inspection and cross-correlation of the temperatures (Fig. 2d), the features appear to be stationary or moving slowly relative to the interior, although they are embedded in large zonal wind currents. Quasi-stationary wave-like features in the tropospheres of both Jupiter and Saturn had been identified in earlier Voyager IRIS^{18,19} and ground-based^{20–22} measurements. Their origin remains speculative. Forcing by a disturbance deeply seated in Jupiter's atmosphere, where the zonal currents are small, is one possibility¹⁸. Alternatively, the features could result from a local instability at the visible cloud level at latitudes where the zonal winds are small¹⁹.

The stratospheric thermal perturbations (Fig. 2e–g) also exhibit regularly spaced zonal features, but they are less confined in latitude, and some move. Earlier ground-based observations had provided evidence of features in the stratosphere, but could only set upper limits to horizontal motions, because of limited temporal sampling³. The CIRS data are able to resolve motions, and they indicate that the temperature features display a systematic westward drift at several latitudes. For example, the warm spot near 25° S, 360° W in Fig. 2e is observed to move westward at a speed of about 50 m s^{-1} . Another isolated warm feature initially at 35° N, 60° W also drifts westward at a comparable rate. Our zonal-wind velocities derived from the thermal-wind equation at 1 mbar (Fig. 1b) show very low wind speeds at 25° S and eastward winds of $\sim 20 \text{ m s}^{-1}$ at 35° N, quite different from the observed motion of the thermal features. Indeed, the cross-correlation curves in Fig. 2h indicate that the thermal features exhibit westward motion relative to the zonal flow through much of the southern hemisphere, from near the equator to 50° , suggesting they may be waves propagating westward. This motion is consistent with planetary, or Rossby, waves¹⁴, but the exact nature of these waves has yet to be determined. Some of the observed stratospheric features are probably propagating disturbances from the troposphere. For example, the 1-mbar features at 10 – 30° N may be related to the long train of features near 15° N seen at 243 mbar. The origin of the westward-moving stratospheric features, however, is not yet evident. To first order, a vertically propagating wave should conserve its zonal phase velocity, so one would expect to see some westward-moving features in the troposphere, which are not visible on casual inspection. However, a more systematic search of the propagation of individual spectral components is needed, like that done for the Voyager thermal infrared data¹⁹.

Prominent in the 1-mbar map in Fig. 2e–g is a warm region at high northern latitudes, centred near 65° N, 180° W. (Although largely excluded from the Fig. 2 maps, there is also a less intense warm region, centred at 77° S, 30° W.) This is better viewed in the north polar projections of Fig. 3. The northern auroral spot is consistent with the locus of enhanced thermal-infrared emission that has been seen in ground-based observations for several years^{6–8}, and appears fixed in latitude and (System III) longitude. It also coincides with a region of excess, pulsating X-ray emission⁹ and anomalous ultraviolet emission in the far ultraviolet^{23,24}. The observed excess infrared emission places the most stringent requirement on the energy input: $\sim 4 \times 10^{13} \text{ W}$, and the probable source is precipitating charged particles that heat the atmosphere at the 20 – μbar level or higher, although Joule heating is another possibility⁸. Tracking Jupiter's magnetic field lines from the hot spot indicates an origin in the outer magnetosphere beyond $30 R_J$ (ref. 9). The impact on the neutral atmosphere must be significant. If a dynamical balance can be achieved, the warmer temperatures within the spot imply a pressure high and an associated anticyclonic (clockwise) vortex. Figure 3a suggests a temperature gradient of at least 15 K per

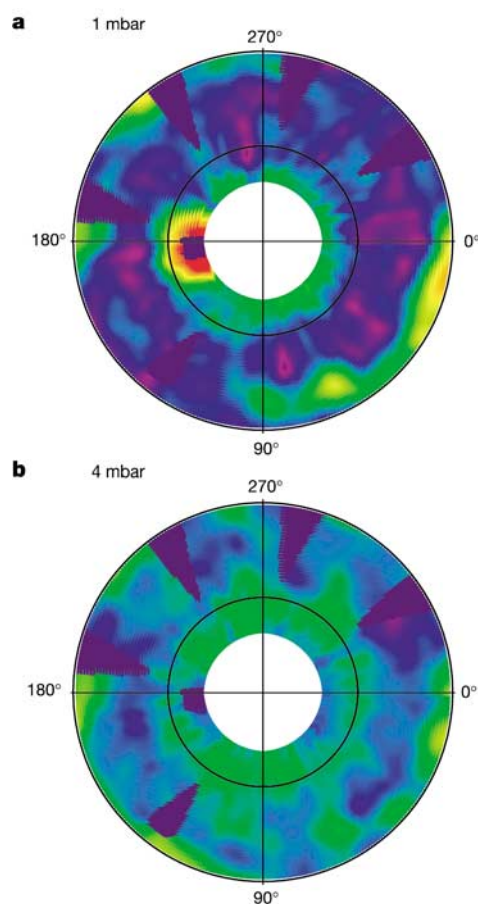


Figure 3 North-polar projection of temperatures shows a hot spot high in the stratosphere. The hot spot is seen at the 1-mbar level (a), but not at the 4-mbar level (b). This is from the 1 January 2001 mapping sequence. Latitude circles are at 60° N and 30° N. In the rainbow colour scale, red to purple span 185 K to 160 K . Although the hot spot is strongly evident in the 1-mbar map, there is an ambiguity as to whether it actually extends down to this level. This is because the emission in the Q-branch of the ν_4 band of CH_4 near $1,304 \text{ cm}^{-1}$, used in the temperature retrievals, can be markedly influenced by hotter temperatures at altitudes up to the 1 – μbar level⁶. Hence, the black region within the hot spot reflects this ambiguity and not a lack of data.

5° of latitude, and the thermal wind equation implies a vertical shear of at least 30 m s^{-1} per scale height ($\sim 27 \text{ km}$) in the vortex winds. A temporal sequence of Cassini ultraviolet images showed a large dark oval in the same general location in late October 2000¹³. The feature exhibited a clockwise rotation, and it subsequently moved eastward and deformed. Its connection with the thermal hot spot remains speculative.

The depth to which the hot spot penetrates has not been well determined, but the CIRS data can place limits. Although it is strongly evident in the 1-mbar map, there is an ambiguity as to whether it actually extends down to this level. This is because the spectral region used to obtain the 1-mbar temperatures can be influenced by thermal emission from higher elevations if they are very warm. However, the hot spot is absent approximately one pressure scale height deeper, at the 4-mbar level (Fig. 3b). This sets an upper bound of 4 mbar to the pressure of the spot. □

Received 8 August; accepted 8 October 2003; doi:10.1038/nature02142.

- Andrews, D. G., Holton, J. R. & Leovy, C. B. *Middle Atmosphere Dynamics* (Academic, New York, 1987).
- Baldwin, M. P. *et al.* The Quasi-Biennial Oscillation. *Rev. Geophys.* **39**, 179–229 (2001).
- Orton, G. S. *et al.* Thermal maps of Jupiter: Spatial organization and time dependence of stratospheric temperatures, 1980 to 1990. *Science* **252**, 537–542 (1991).
- Leovy, C. B., Friedson, A. J. & Orton, G. S. The quasiquadrennial oscillation of Jupiter's equatorial stratosphere. *Nature* **354**, 380–382 (1991).
- Friedson, A. J. New observations and modeling of a QBO-like oscillation in Jupiter's stratosphere. *Icarus* **137**, 34–55 (1999).
- Caldwell, J., Tokunaga, A. T. & Gillett, F. C. Possible infrared aurorae on Jupiter. *Icarus* **44**, 667–675 (1980).
- Caldwell, J. J., Halthore, R., Orton, G. & Bergstralh, J. Infrared polar brightenings on Jupiter. 4. Spatial properties of methane emission. *Icarus* **74**, 331–339 (1988).
- Drossart, P. *et al.* Thermal profiles in the auroral regions of Jupiter. *J. Geophys. Res.* **98**, 18803–18811 (1993).
- Gladstone, G. R. *et al.* A pulsating auroral X-ray hot spot on Jupiter. *Nature* **415**, 1000–1003 (2002).
- Smith, B. A. *et al.* The Galilean satellites and Jupiter: Voyager 2 imaging science results. *Science* **206**, 927–950 (1979).
- Limaye, S. S. Jupiter: New estimates of the mean zonal flow at the cloud level. *Icarus* **65**, 335–352 (1986).
- Simon, A. A. The structure and temporal stability of Jupiter's zonal winds: A study of the north tropical region. *Icarus* **141**, 29–39 (1999).
- Porco, C. C. *et al.* Cassini imaging of Jupiter's atmosphere, satellites, and rings. *Science* **299**, 1541–1547 (2003).
- Holton, J. R. *An Introduction to Dynamic Meteorology* 3rd edn (Academic, New York, 1992).
- Kunde, V. G. *et al.* Cassini infrared Fourier spectroscopic investigation. *Proc. SPIE* **2803**, 162–177 (1996).
- Gierasch, P. J., Conrath, B. J. & Magalhães, J. A. Zonal mean properties of Jupiter's upper troposphere from Voyager infrared observations. *Icarus* **67**, 456–483 (1986).
- Li, X. & Read, P. L. A mechanistic model of the quasi-quadrennial oscillation in Jupiter's stratosphere. *Planet. Space Sci.* **48**, 637–669 (2000).
- Magalhães, J. A. *et al.* Zonal motion and structure in Jupiter's upper troposphere from Voyager infrared and imaging observations. *Icarus* **89**, 39–72 (1990).
- Achterberg, R. K. & Flasar, F. M. Planetary-scale thermal waves in Saturn's upper troposphere. *Icarus* **119**, 350–369 (1996).
- Deming, D. *et al.* A search for p-mode oscillations of Jupiter: Serendipitous observations of nonacoustic thermal wave structure. *Astrophys. J.* **343**, 456–467 (1989).
- Deming, D. *et al.* Observations and analysis of longitudinal thermal waves on Jupiter. *Icarus* **126**, 301–312 (1997).
- Orton, G. S. *et al.* Spatial organization and time dependence of Jupiter's tropospheric temperatures 1980–1993. *Science* **265**, 625–631 (1994).
- Pallier, L. & Prangé, R. More about the structure of the high latitude Jovian aurorae. *Planet. Space Sci.* **49**, 1159–1173 (2001).
- Waite, J. H. *et al.* An auroral flare at Jupiter. *Nature* **410**, 787–789 (2001).
- Beebe, R. *Jupiter* (Smithsonian Institution Press, Washington, 1994).
- Seidelmann, P. K. *et al.* Report of the IAU/IAG working group on cartographic coordinates and rotational elements of the planets and satellites: 2000. *Celest. Mech. Dynam. Astron.* **82**, 83–110 (2002).
- Conrath, B. J., Gierasch, P. J. & Ustinov, E. A. Thermal structure and para hydrogen on the outer planets from Voyager IRIS measurements. *Icarus* **135**, 501–517 (1998).
- Riddle, A. C. & Warwick, J. W. Redefinition of System III longitude. *Icarus* **27**, 457–459 (1976).

Acknowledgements The contributions of T.F. to this work were made while he was at AOP, Oxford University. We thank M. H. Elliott, J. S. Tingley, F. Carroll and M. E. Segura for assistance with instrument commanding and data processing; P. J. Schinder for computing the pointing geometry files for the CIRS observations during the Jupiter swing-by; and D. L. Matson for suggestions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.M.F. (f.m.flasar@nasa.gov).

Hybridization of electronic states in quantum dots through photon emission

Khaled Karrai¹, Richard J. Warburton², Christian Schulhauser¹, Alexander Högele¹, Bernhard Urbaszek², Ewan J. McGehee², Alexander O. Govorov^{3,4}, Jorge M. Garcia⁵, Brian D. Gerardot⁶ & Pierre M. Petroff⁶

¹Center for NanoScience and Sektion Physik, Ludwig-Maximilians-Universität, Geschwister-Scholl-Platz 1, 80539 München, Germany

²School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

³Department of Physics and Astronomy, Ohio University, Athens, Ohio USA

⁴Institute of Semiconductor Physics, Russian Academy of Sciences, Siberian Branch, 630090 Novosibirsk, Russia

⁵Instituto de Microelectronica de Madrid, CNM-CSIC Isaac Newton, 8, PTM, 28760 Madrid, Spain

⁶Materials Department, University of California, Santa Barbara, California 93106, USA

The self-assembly of semiconductor quantum dots has opened up new opportunities in photonics. Quantum dots are usually described as ‘artificial atoms’, because electron and hole confinement gives rise to discrete energy levels. This picture can be justified from the shell structure observed as a quantum dot is filled either with excitons¹ (bound electron–hole pairs) or with electrons². The discrete energy levels have been most spectacularly exploited in single photon sources that use a single quantum dot as emitter^{3–6}. At low temperatures, the artificial atom picture is strengthened by the long coherence times of excitons in quantum dots^{7–9}, motivating the application of quantum dots in quantum optics and quantum information processing. In this context, excitons in quantum dots have already been manipulated coherently^{10–12}. We show here that quantum dots can also possess electronic states that go far beyond the artificial atom model. These states are a coherent hybridization of localized quantum dot states and extended continuum states: they have no analogue in atomic physics. The states are generated by the emission of a photon from a quantum dot. We show how a new version of the Anderson model that describes interactions between localized and extended states can account for the observed hybridization.

The starting point for our experiments is the generation of excitons with an unambiguous charge in individual self-assembled quantum dots² (see also Supplementary Information). Figure 1 shows the magnetic dispersions of the photoluminescence (PL) from the singly-, doubly- and triply-charged excitons (X^{1-} , X^{2-} and X^{3-} , respectively), all on the same dot. For X^{1-} and X^{2-} , the behaviour is typical of localized excitons. There is a diamagnetic shift arising from the enhancement of the exciton confinement, and a splitting arising from the spin Zeeman effect¹³. However, X^{3-} has a completely different behaviour. At high field, in addition to the spin Zeeman effect, the PL develops a remarkable series of oscillations. This is preceded by a gradual collapse around 1 T of the two low field lines. As we show, the collapse can be explained with the artificial atom model, but the series of oscillations can only be explained with a new electronic interaction.

The X^{3-} PL is represented in Fig. 2 by plotting the energy of each PL peak against magnetic field. Clearly, there is a series of anti-crossings. A number of factors point to an interaction with Landau levels. (Landau levels have energies $E_n = (n + \frac{1}{2})\hbar\omega_c$, where $n = 0, 1, 2, \dots$, and arise when the kinetic energy in an electron band is quantized into units of the cyclotron energy, $\hbar\omega_c$, by a magnetic field.) First, the asymptotes at each anti-crossing are linear

functions of magnetic field, B . Second, the asymptotes for all the anti-crossings extrapolate back to the same energy at zero magnetic field. Third, the PL intensity at each energy is a periodic function of $1/B$. All these features are characteristic of Landau levels in a two-dimensional system. The surprising result is that the PL from an exciton bound to a small quantum dot takes on the

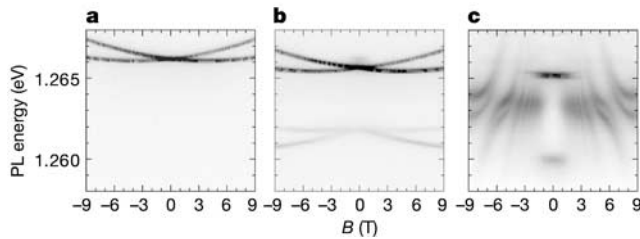


Figure 1 Photoluminescence from charged excitons. Data are shown for singly (a), doubly (b) and triply (c) charged excitons, all from the same quantum dot, versus magnetic field. The photoluminescence (PL) is represented with a grey-scale, and the magnetic field is applied perpendicular to the quantum dot plane. The InAs quantum dots are grown in the Stranski–Krastanow mode on GaAs with an annealing step at the growth temperature in order to increase the ensemble PL energy from 1.1 to 1.3 eV. Under special conditions, the annealing step can lead to the formation of quantum rings. Here, we use the annealing step to form heavily alloyed dots with a vertical height of about 3 nm and a lateral extent of about 40 nm. The heterostructure sequence is GaAs (buffer layer), n^+ -GaAs (back contact), 25 nm GaAs (tunnel barrier), quantum dots, and 150 nm GaAs/AlAs superlattice (blocking barrier). Contacts are made to the n^+ layer and a 5 nm NiCr layer (gate electrode) is deposited on the surface. A voltage V_g is applied between gate and back contact. The sample is mounted into a miniature confocal microscope and cooled to 4.2 K. The PL is excited by non-resonant excitation of the wetting layer with a 830-nm laser diode; the PL is dispersed with a grating spectrometer and detected with a silicon charged-coupled-device camera (spectral resolution 0.1 meV). The excitation power is low enough that any biexciton-related features are much weaker than the exciton PL. The exciton charge can be set unambiguously with the gate voltage because of the strong Coulomb blockade (see Supplementary Information). The neutral exciton is not shown because the shift and splitting in magnetic field are identical to those of the singly charged exciton. The X^{2-} PL has two lines as a consequence of exchange-split final states, one with electron spin $S = 1$, the other with $S = 0$ (ref. 2).

properties of extended states at much higher energies, the Landau levels.

The anti-crossings prove that the quantum dot interacts coherently with the Landau levels. From the splitting at the 6.5-T resonance (2.4 meV), we can deduce that the system oscillates between localized quantum dot character and Landau level character at a frequency of 0.6 THz. The asymptotes shown in Fig. 2 have a negative gradient. Bearing in mind that a PL process involves an initial and a final state, the negative dispersion demonstrates that the hybridization takes place in the final state. In other words, photon emission triggers the hybridization. A resonant interaction with a phonon, a feature in the PL of doped two-dimensional systems¹⁴, can be ruled out because such an interaction would be common to both X^{2-} and X^{3-} , yet in our case, the interaction is completely absent for X^{2-} . We present an explanation of the hybridization based on a purely electronic interaction.

It is important to establish the relative magnitude of the quantization energy and a typical Coulomb energy for the dot studied here. The X^0 diamagnetic shift is both small ($10 \mu\text{eV T}^{-2}$), signifying a strongly confined exciton, and unchanged on charging. This is characteristic of the strong confinement regime where the quantization energy is larger than a typical Coulomb energy¹³. (When the quantization and Coulomb energies are comparable, the diamagnetic shift is larger and decreases significantly on charging¹³.) We reach the same conclusion from the Coulomb blockade. The X^{1-} extends over a much larger gate voltage range than the X^0 and X^{2-} plateaus (see Supplementary Information), demonstrating the large quantization energy. We therefore describe the charged excitons by treating the Coulomb interactions as perturbations to the single-particle states. The dot has s and p orbitals, separated by about 34 meV as deduced from the diamagnetic shift, and we know that the two p orbitals are close to degenerate from the form of the X^{3-} PL at $B = 0$ (explained below). Finally, the d orbitals cannot exist as the X^{3-} is the most highly charged exciton that we can generate before charge occupies not the dot but the wetting layer.

Within the artificial atom model, the X^{3-} emission process is displayed in Fig. 3. The initial state has electron spin $S = 1$ through Hund's rule. When a photon is emitted, there are two possible final

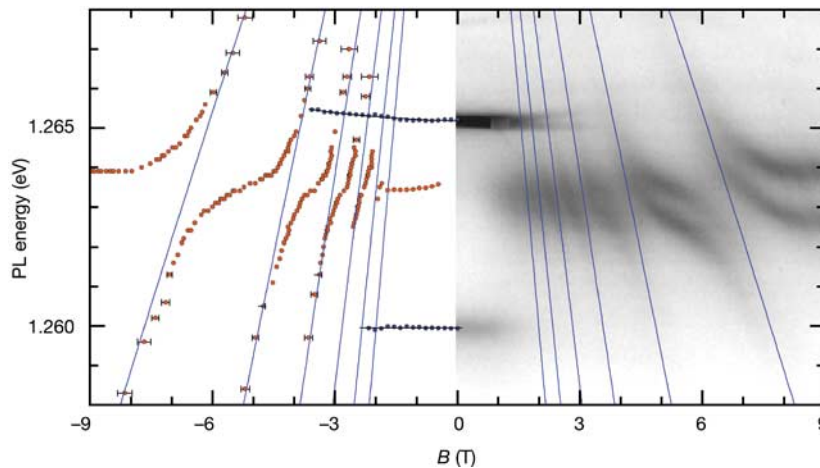


Figure 2 Photoluminescence energy versus magnetic field for the X^{3-} exciton. The points in the plot on the left correspond to the energies of the PL lines shown in Fig. 1c and reproduced on the right. For simplicity, just one spin branch has been included in the plot on the left; the other spin branch shows the same behaviour but shifted in energy. The solid lines show the dependence of energy $E_0 + \Delta - (n + \frac{3}{2})e\hbar B/m_0^*$ on magnetic field B with $E_0 = 1.2635 \text{ eV}$, $\Delta = 20 \text{ meV}$ and $m^* = 0.07 - 0.0018|B| + 0.00005B^2$

where Δ is the excess kinetic energy, n the Landau level index, and m^* the effective mass in units of the free electron mass, m_0 . These lines were determined by fitting this functional form to the asymptotes of the anti-crossings. These asymptotes are also shown on top of the original data (right). We note that inclusion of the diamagnetic shift in the analysis does not significantly change the masses that we deduce.

states, $S = 3/2$ and $S = 1/2$, energetically distinct through electron exchange. This is the origin of the two lines in the X^{3-} emission at $B = 0$, the higher- (lower-) energy line corresponding to emission into the $S = 3/2$ ($S = 1/2$) state. A magnetic field introduces an orbital Zeeman splitting between the two p orbitals¹⁵, such that when the splitting exceeds the exchange energy, it becomes energetically favourable for both p electrons to occupy the lower energy p orbital and form an $S = 0$ state¹⁶. We attribute the collapse of the PL splitting around 1 T to this change in configuration. There are two compelling arguments. First, this field can be anticipated from the splitting in the PL at zero field, ΔE . Assuming an isotropic harmonic confinement¹⁷, the configuration change occurs when $\hbar\omega_c = \frac{1}{4}\Delta E$, implying a magnetic field of 0.8 T. This is in good agreement with the experiment where the $S = 1$ and $S = 0$ X^{3-} PL lines have equal intensities at about 1.2 T. Second, once the X^{3-} forms an $S = 0$ state, there is only one possible final state energy, and it is therefore natural for the PL to lie between the two exchange-split lines of the $S = 1$ X^{3-} , exactly as observed. However, within this artificial atom picture, there is no explanation for the anti-crossings that develop at higher magnetic field.

In atomic physics, configurations with the same good quantum numbers and similar energies are admixed. For the $S = 1$ X^{3-} , and also the X^{2-} , the final state configurations involve only the s and p orbitals. However, we find a new possibility for the final state of the $S = 0$ X^{3-} , which has a doubly occupied p orbital and a singly occupied s orbital. We consider first a hypothetical quantum dot with a bound d orbital. The quantum numbers can be preserved by allowing one of the p electrons to fill the vacancy in the s orbital, with the other p electron occupying a higher energy d orbital. These two configurations (Fig. 4a) have the same values of angular momentum and spin and are also almost degenerate, and are therefore admixed. For our dot, the d orbital is unbound so that the mechanism now admixes continuum states to the quantum dot wavefunction (Fig. 4b). The continuum states are associated with the two-dimensional wetting layer. In a magnetic field, the

continuum states are quantized into Landau levels, such that the final state becomes a hybridization of quantum dot states and Landau levels. This means that continuum states impart their character to the PL even though they are never occupied in the initial state. This hybridization occurs for the $S = 0$ X^{3-} but not for the $S = 1$ X^{3-} or the more weakly charged excitons, and there is therefore a close correspondence with the experiments, where the $S = 0$ X^{3-} interacts with the Landau levels but the other excitons do not. Additionally, the interaction takes place in the final state, as required.

The exciton final state is strongly hybridized whenever the energy between the occupied p orbital and the s orbital exactly matches the energy between a Landau level and the p orbital (Fig. 4c). Hybridization with the n th Landau level occurs when $\Delta = (n + \frac{3}{2})\hbar\omega_c$ where Δ is the excess kinetic energy, defined in Fig. 4b ($\Delta = 20$ meV for the dot in Fig. 1). (We have assumed that the p orbital moves down in energy by $\frac{1}{2}\hbar\omega_c$, an exact result for an isotropic and harmonic confinement¹⁵.) As the magnetic field is increased, this condition is satisfied for each Landau level in turn, resulting in the periodicity in $1/B$. Note that hybridization with every Landau level is allowed because every Landau level has a d -like component independent of the Landau level index¹⁸. Ultimately, at high magnetic field (8 T for our particular dot), the interaction with the Landau levels is suppressed because the separation between the s and p orbitals becomes insufficient.

To confirm that our mechanism accounts quantitatively for the experimental results, we have calculated the matrix element $M_n = \langle s^\uparrow, p^\uparrow, p^\downarrow | V_C | s^\uparrow, s^\downarrow, n^\uparrow \rangle$ where $|s^\uparrow, p^\uparrow, p^\downarrow\rangle$ denotes the configuration with two electrons in the lower p orbital and one electron in the s orbital, and $|s^\uparrow, s^\downarrow, n^\uparrow\rangle$ denotes the configuration with two electrons in the s orbital and one in the n th Landau level. V_C is the Coulomb interaction, and $\uparrow$ ($\downarrow$) denotes spin-up (spin-down). With harmonic oscillator wavefunctions, and Landau level wavefunctions in the

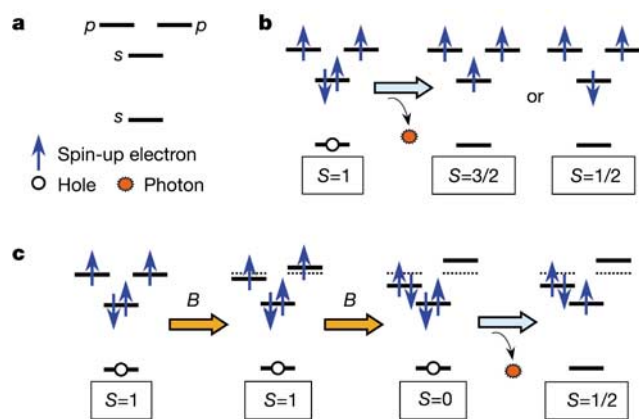


Figure 3 Quantum dot level diagrams in the artificial atom model. **a**, Definition of the orbitals and symbols. Only the lowest-energy hole orbital, labelled s , is shown, as only this hole orbital participates in the recombination. The electron orbitals are labelled s and p with z -component of angular momentum $m = 0, \pm 1$, respectively. Electrons are shown as $\uparrow$ ($\downarrow$) according to their z -component of spin; the hole has spin $\pm 3/2$. **b**, Configurations are shown for the initial and final states of the triply charged exciton X^{3-} . S refers to the total electron spin in each case. The initial state has $S = 1$. The final state configurations are shown in a simplified way. (The $|S, S_z\rangle = |3/2, 1/2\rangle$ and $|S, S_z\rangle = |1/2, 1/2\rangle$ final states are admixtures of the three configurations in each of which one electron has spin down.) **c**, The initial state configuration change from $S = 1$ to $S = 0$ is shown, induced by a magnetic field.

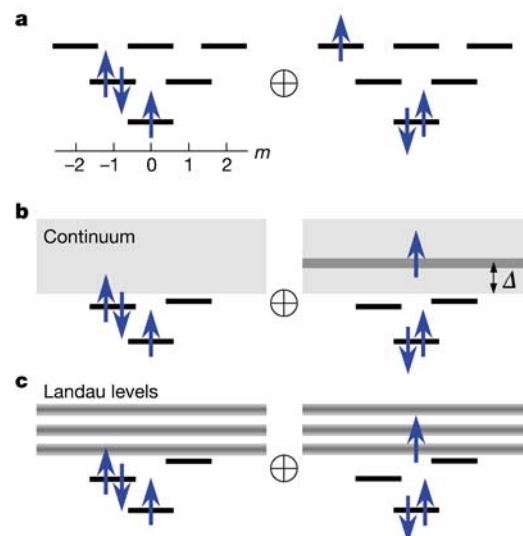


Figure 4 Hybridization in the final state after photon emission. **a**, The artificial atom model including the d orbitals with angular momentum $-2, 0$ and $+2$. The configuration on the left is the final state of the $S = 0$ X^{3-} exciton. It is admixed with the configuration on the right, which has the same spin and angular momentum quantum numbers and almost the same energy. **b**, The admixture when the d orbitals do not exist and there is instead a continuum associated with the two-dimensional wetting layer. The admixture imparts continuum character to the quantum dot exciton. **c**, A large magnetic field is applied to condense the continuum states into Landau levels. The situation shown corresponds to $B = 6.5$ T in the experiment: there is a hybridization of the final state with the $n = 0$ Landau level.

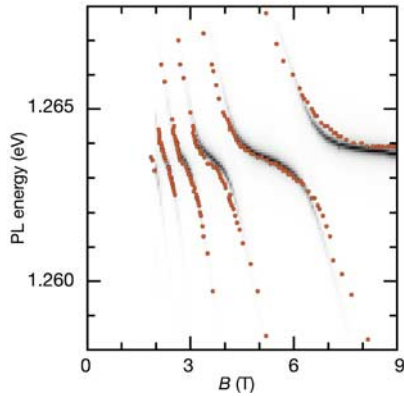


Figure 5 A comparison of the results from the Anderson hamiltonian model with the experiment for the X^{3-} exciton. The grey-scale shows the results of the calculation with the intensity representing the oscillator strength; the points are the same as in Fig. 2 and represent the results of the experiment. Results are plotted with $E_0 = 1.2635$ eV, $\Delta = 20$ meV, $M = 1.2$ meV and $m^* = 0.07 - 0.0018|B| + 0.00005B^2$.

symmetric gauge¹⁸ centred on the dot

$$M_0 = \frac{e^2}{4\pi\epsilon_0\epsilon_r} \frac{16\sqrt{\pi}L_c^3L_c^2(3L_c^2 + L_c^2)}{(2L_c^2 + L_c^2)^{3/2}(6L_c^2 + L_c^2)^{5/2}} \quad (1)$$

where L_c is the characteristic length of the electron wavefunction¹⁷ and $L_c = \sqrt{\hbar/eB}$ is the magnetic length. We determine L_c to be 5.6 nm from the diamagnetic shift of the X^0 PL. At 6.5 T, the field for the resonance with the $n = 0$ Landau level, equation (1) predicts a splitting between the PL lines of $2|M_0| = 2.0$ meV, in very good agreement with the experimental result, 2.4 meV.

The presence of continuum states results in absorption features close in energy to high-energy quantum dot transitions, and because of this the continuum can dephase excited excitons^{19–21}. The mechanism we propose here is significantly different not only because the emission arises from recombination between the lowest-energy electron and hole orbitals, but also because the interaction with the continuum in magnetic field is not a source of scattering but is coherent. We anticipate that the hybridization is by no means unique to our particular quantum dots. The interaction will be a feature for all more highly charged excitons, X^{4-} , X^{5-} , X^{6-} , and so on, because in each case after photon emission, there is a vacancy in the s orbital with a filled p orbital, the necessary requirements for the hybridization to take place. For very deep dots where the d orbital is bound, the interaction will admix d character, but ultimately—for the most highly charged excitons such a dot can support—the exciton energy will lie just below the continuum such that a hybridization with the continuum takes place.

The physics that we are describing is a novel interaction between a localized state and a continuum at higher energy. Such problems are often described by the Anderson hamiltonian²². We have developed a new version of the Anderson model to describe the emission properties of the $S = 0$ X^{3-} . The beauty of this approach is that it relates the optical spectrum, $I(E)$, directly to the density of states in the continuum

$$I(E) \propto \text{Re} \frac{-i}{E - \Sigma(E) - i0}$$

where the self-energy Σ is given by

$$\Sigma(E) = \int_0^\infty \frac{|M(E')|^2 \rho(E') dE'}{E - \Delta + E' - i0}$$

where $M(E)$ is the Coulomb matrix element and ρ the density of

states. Noting that $M(E)$ decreases with increasing energy, we simplify the calculation of Σ by making the approximation that $M(E) = M$, which is a constant for $0 \leq E \leq \hbar^2/(m^*m_0L_c^2)$, and zero at higher energy. M can be taken from equation (1). However, the theory offers an additional opportunity as we discover that M is related to the linewidth of the $S = 0$ X^{3-} PL at low magnetic field, Γ , and the cyclotron energy at the resonance with the $n = 0$ Landau level, $\hbar\omega_c^0$, through $M = \sqrt{\Gamma\hbar\omega_c^0/\pi}$. The low-field linewidth is $\Gamma = 0.40$ meV, implying $2M = 2.4$ meV, in excellent agreement with the experiment. The results of the calculation, using the experimentally deduced effective mass (Fig. 2), are shown in Fig. 5 and give excellent agreement with the measured X^{3-} PL. This demonstrates conclusively that the hybridization in the experiment is caused by an interaction between localized and extended states, as this is the main content of the Anderson hamiltonian. A natural question is what happens when the wetting layer is filled with electrons. In this case, we find that the $S = 0$ X^{3-} state is favoured even at zero magnetic field (see Supplementary Information), and we anticipate that our new electronic hybridization can coherently couple the quantum dot with the Fermi sea, leading to Kondo-like effects in the emission. □

Received 23 July; accepted 2 October 2003; doi:10.1038/nature02109.

1. Bayer, M., Stern, O., Hawrylak, P., Fafard, S. & Forchel, A. Hidden symmetries in the energy levels of excitonic 'artificial atoms'. *Nature* **405**, 923–926 (2000).
2. Warburton, R. J. *et al.* Optical emission from a charge-tunable quantum ring. *Nature* **405**, 926–929 (2000).
3. Michler, P. *et al.* A quantum dot single-photon turnstile device. *Science* **290**, 2282–2285 (2000).
4. Moreau, E. *et al.* Quantum cascade of photons in semiconductor quantum dots. *Phys. Rev. Lett.* **87**, 183601 (2001).
5. Yuan, Z. *et al.* Electrically driven single-photon source. *Science* **295**, 102–105 (2001).
6. Santori, C., Fattal, D., Vučković, J., Solomon, G. S. & Yamamoto, Y. Indistinguishable photons from a single-photon device. *Nature* **419**, 594–597 (2002).
7. Gammon, D., Snow, E. S., Shanabrook, B. V., Katzer, D. S. & Park, D. Homogeneous linewidths in the optical spectrum of a single gallium arsenide quantum dot. *Science* **273**, 87–90 (1996).
8. Borri, P. *et al.* Ultralong dephasing time in InGaAs quantum dots. *Phys. Rev. Lett.* **87**, 157401 (2001).
9. Bayer, M. & Forchel, A. Temperature dependence of the exciton homogeneous linewidth in $\text{In}_{0.46}\text{Ga}_{0.54}\text{As}$ /GaAs self-assembled quantum dots. *Phys. Rev. B* **65**, 041308 (2002).
10. Bonadeo, N. H. *et al.* Coherent optical control of the quantum state of a single quantum dot. *Science* **282**, 1473–1476 (1998).
11. Kamada, H., Gotoh, H., Temmyo, J., Takagahara, T. & Ando, H. Exciton Rabi oscillation in a single quantum dot. *Phys. Rev. Lett.* **87**, 246401 (2001).
12. Zrenner, A. *et al.* Coherent properties of a two-level system based on a quantum-dot photodiode. *Nature* **418**, 612–614 (2002).
13. Schulhauser, C. *et al.* Magneto-optical properties of charged excitons in quantum dots. *Phys. Rev. B* **66**, 193303 (2002).
14. Simmonds, P. E., Skolnick, M. S., Fisher, T. A., Nash, K. J. & Smith, R. S. Hole-quasiparticle resonant polaron coupling in quantum wells containing high densities of free carriers. *Phys. Rev. B* **45**, 9497–9500 (1992).
15. Fock, V. Bemerkung zur Quantelung des harmonischen Oszillators im Magnetfeld. *Z. Phys.* **47**, 446–448 (1928).
16. Tarucha, S., Austing, D. G., Honda, T., van der Hage, R. J. & Kouwenhoven, L. P. Shell filling and spin effects in a few electron quantum dot. *Phys. Rev. Lett.* **77**, 3613–3616 (1996).
17. Warburton, R. J. *et al.* Coulomb interactions in small, charge-tunable quantum dots: A simple model. *Phys. Rev. B* **58**, 16221–16231 (1998).
18. Dingle, R. B. Some magnetic properties of metals I. General introduction, and properties of large systems of electrons. *Proc. R. Soc. Lond. A* **211**, 500–516 (1951).
19. Toda, Y., Moriawaki, O., Nishioka, M. & Arakawa, Y. Efficient carrier relaxation mechanism in InGaAs/GaAs self-assembled quantum dots based on the existence of continuum states. *Phys. Rev. Lett.* **82**, 4114–4117 (1999).
20. Kammerer, C. *et al.* Photoluminescence up-conversion in single self-assembled InAs/GaAs quantum dots. *Phys. Rev. Lett.* **87**, 207401 (2001).
21. Vasanelli, A., Ferreira, R. & Bastard, G. Continuous absorption background and decoherence in quantum dots. *Phys. Rev. Lett.* **89**, 216804 (2002).
22. Hewson, A. C. *The Kondo Problem to Heavy Fermions* (Cambridge Univ. Press, Cambridge, 1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We acknowledge discussions with A. Rosch and J. von Delft. This work was funded by the DFG, EPSRC and The Royal Society.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.J.W. (R.J.Warburton@hw.ac.uk).

Detection of molecular interactions at membrane surfaces through colloid phase transitions

Michael M. Baksh^{1,2}, Michal Jaros¹ & Jay T. Groves^{1,2}

¹Department of Chemistry, University of California and ²Physical Bioscience Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

The molecular architecture of—and biochemical processes within—cell membranes play important roles in all living organisms, with many drugs and infectious disease agents targeting membranes. Experimental studies of biochemical reactions on membrane surfaces are challenging, as they require a membrane environment that is fluid (like cell membranes) but nevertheless allows for the efficient detection and characterization of molecular interactions. One approach uses lipid membranes supported on solid substrates such as silica or polymers^{1,2}; although the membrane is trapped near the solid interface, it retains natural fluidity and biological functionality³ and can be implanted with membrane proteins for functional studies⁴. But the detection of molecular interactions involving membrane-bound species generally requires elaborate techniques, such as surface plasmon resonance⁵ or total internal reflection fluorescence microscopy⁶. Here we demonstrate that colloidal phase transitions of membrane-coated silica beads provide a simple and label-free method for monitoring molecular interactions on lipid membrane surfaces. By adjusting the lipid membrane composition and hence the pair interaction potential between the membrane-supporting silica beads, we poise our system near a phase transition so that small perturbations on the membrane surface induce dramatic changes in the macroscopic organization of the colloid. We expect that this approach, used here to probe with high sensitivity

protein binding events at membrane surfaces, can be applied to study a broad range of cell membrane processes.

Colloidal particles self-assemble into a variety of ordered phases, and the study of colloidal phase transitions is currently an area of intensive activity⁷. This collective effort is fuelled by the intriguing statistical thermodynamic behaviour of colloids, their general applicability to the study of phase transitions, and their important material properties, such as the ability to self-assemble into ordered structures^{8–12}. The behaviour of a colloidal system is driven by the pair interaction potential between particles. In the case of membrane-derivatized silica beads, the pair potential is dominated by membrane–membrane interactions. Two-dimensional dispersions of lipid-membrane-derivatized silica beads exhibit colloidal phase transitions that are governed by details of these membrane surface interactions. The collective phase behaviour serves as a cooperative amplifier that produces a readily detectable response from a small number of molecular events on the membrane surface. Using direct optical imaging, we observe multiple near-equilibrium phases and find that protein binding to membrane-associated ligands at densities as low as 10^{-4} monolayer can trigger a phase transition. Statistical analysis of bead pair distribution functions enables quantitative comparison among different membrane systems and reveals subtle, pre-transition effects.

Lipid membranes can be assembled on silica beads by essentially the same vesicle fusion process used to form supported membranes on monolithic substrates^{13,14}. The resulting membrane, illustrated schematically in Fig. 1a, is continuous and retains lateral fluidity. Figure 1b and c depicts fluorescence recovery after photobleaching (FRAP) experiments performed on beads coated with fluid and

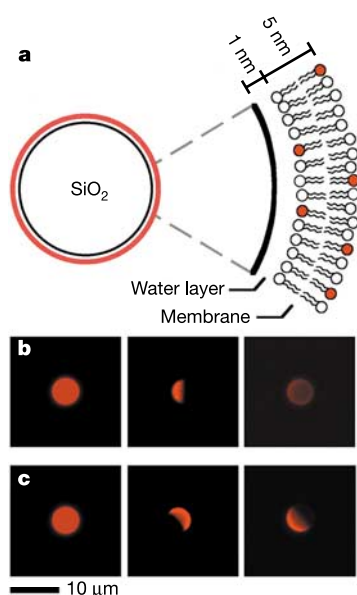


Figure 1 Membrane-derivatized bead. **a**, Schematic illustration of a membrane-derivatized silica bead. **b,c**, FRAP experiments conducted on beads coated with fluid (**b**; 88% 1,2-dimyristoleoyl-*sn*-glycero-3-phosphocholine (DMOPC), 11% 1,2-dioleoyl-*sn*-glycero-3-ethylphosphocholine (DOEPC), 1% *N*-(Texas red sulphonyl)-1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine (Texas red-DPPE) fluorescent probe) and non-fluid (**c**; 88% 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC), 11% DOEPC, 1% Texas red-DPPE) lipid membranes. Full illumination before bleach (left panel), exposure pattern during bleach (middle panel), full illumination 1 min after bleach (right panel).

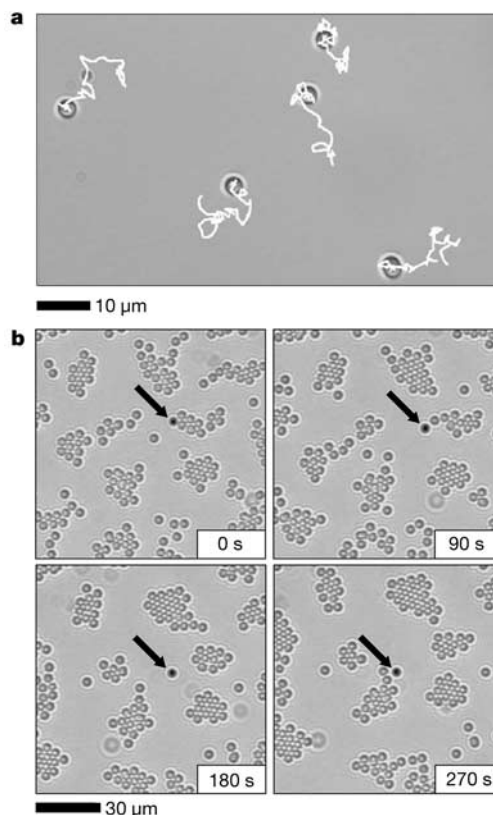


Figure 2 Mobility of membrane-derivatized beads. **a**, Two-dimensional brownian trajectories of membrane-derivatized beads, which have settled gravitationally to the bottom of a dish filled with water. Diffusion coefficients are independent of membrane composition, and ranged from 0.079 to $0.086 \mu\text{m}^2 \text{s}^{-1}$ for $5\text{-}\mu\text{m}$ -diameter beads. **b**, Time sequence images of a condensed phase of the colloid, illustrating the mobility of individual beads into and out of condensed crystallites. The arrow identifies the moving position of an individual bead through the four frames.

non-fluid (gel state) membranes, respectively. Diffusion coefficients of $1\text{--}5\ \mu\text{m}^2\text{s}^{-1}$, as seen here, are typical of fluid membranes. Silica beads can be derivatized with membranes varying widely in composition, providing a precise method of adjusting the chemical and biological constitution of the surface.

Membrane-derivatized silica beads were dispersed, underwater, where they settle gravitationally onto the underlying substrate and form a two-dimensional colloid. The beads exhibit free lateral diffusion, and the system behaves as an ergodic fluid. Brownian trajectories for a dilute collection of beads are illustrated in Fig. 2a. Bead diffusion coefficients are essentially independent of membrane composition; measurements ranged from 0.079 to $0.086\ \mu\text{m}^2\text{s}^{-1}$ for $5\text{-}\mu\text{m}$ -diameter beads. These values are $\sim 80\%$ of that predicted by the Stokes–Einstein relation for purely viscous drag, indicating a small contribution from drag on the underlying substrate. Depending on the strength of the interaction between membranes on the bead surfaces, dispersed (gas) or condensed (liquid or crystalline) phases of the colloid are observed. Bead mobility is retained in condensed phases (Fig. 2b). The mobility of individual beads, in both condensed and dispersed phases, defines the rate of system equilibration. The timescale of bead condensation onto and evaporation from the condensed crystallites, seen in Fig. 2b, is rapid compared to that of our experiments (several minutes versus more than half an hour). Additionally, the overall morphology and quantitative pair distribution functions of the phases remain constant, despite the interchange of individual beads. These observations suggest that the system is near equilibrium, at least over length scales of several bead diameters.

The chemical composition of the lipid membrane was adjusted to modulate the pair interaction potential. Condensed phases, as seen in Fig. 2b, formed whenever the coating membrane was net neutral or negatively charged. In contrast, net positively charged

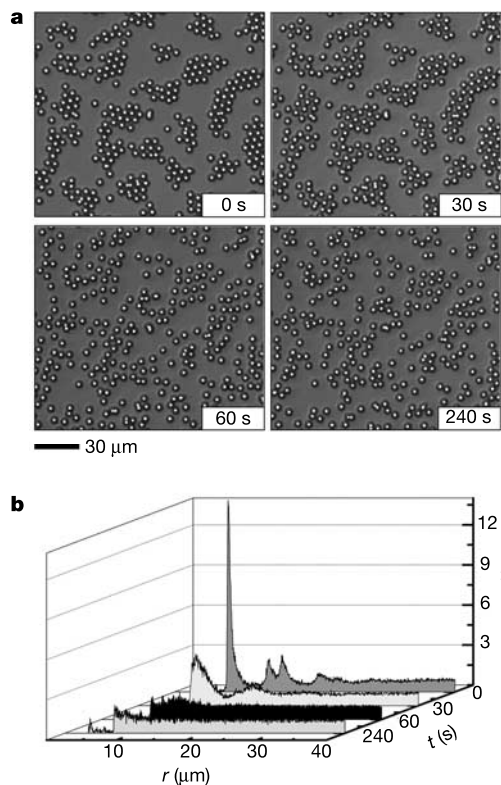


Figure 3 Protein-binding-triggered colloidal phase transition. **a**, Time sequence of images depicting the transition from a condensed to a dispersed colloidal phase, triggered by addition of protein ($5\ \mu\text{g ml}^{-1}$ final concentration anti-Texas red rabbit monoclonal IgG antibody). Transitions were triggered only when the appropriate ligand was also incorporated into the membrane. **b**, Corresponding plots of $g(r)$ for the time sequence in **a**.

membranes led to dispersed phases. The occurrence of multiple phases indicates that pair interaction energies place the system near a phase transition. As such, small perturbations on the membrane surface are expected to induce significant changes in the macroscopic organization of the colloid. We observe this predicted behaviour when examining the effects of protein binding to membrane-associated ligands.

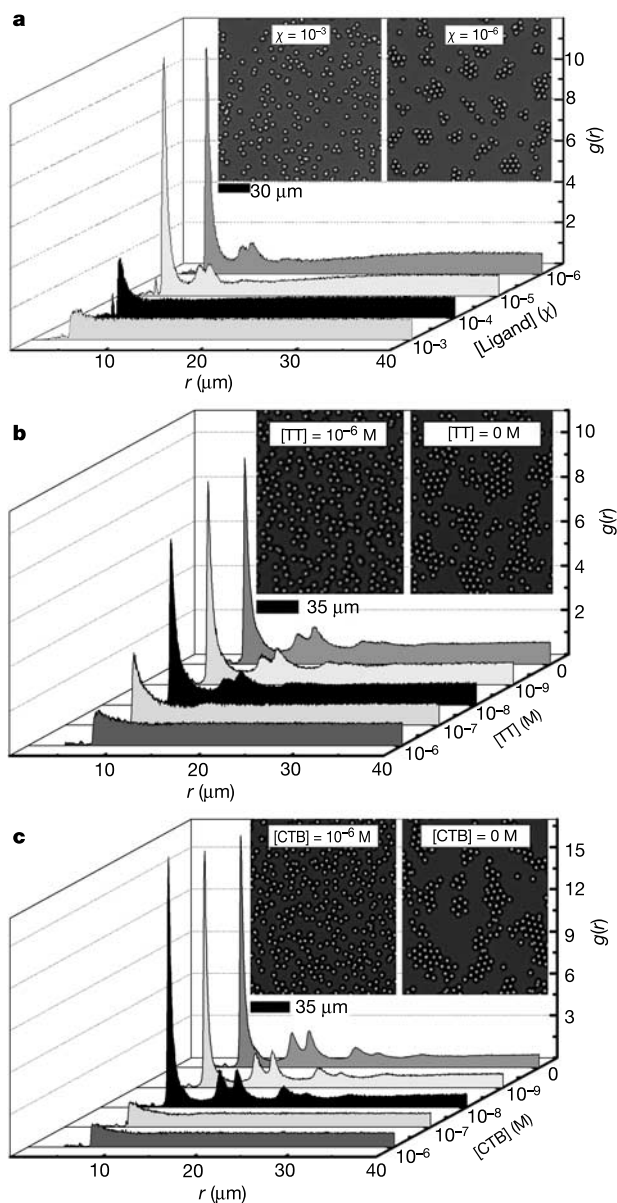


Figure 4 Protein-binding assays. **a**, Plots of measured $g(r)$ functions for dispersions of beads (area fraction $\phi = 0.15$) derivatized with fluid membranes (90% DMOPC, $\sim 9\%$ 1,2-dimyristoyl-*sn*-glycero-3-[phospho-L-serine] (DMPs)) containing different mole fractions (χ) of Texas red–DPPE ligand, after incubation with $20\ \mu\text{g ml}^{-1}$ anti-Texas red rabbit monoclonal IgG antibody. Dispersed phases formed for samples containing the membrane surface ligand at $\chi \geq 10^{-4}$. Inset images depict representative distributions, as labelled. **b**, Plots of $g(r)$ for a series of identical dispersions of $6.8\text{-}\mu\text{m}$ -diameter beads ($\phi = 0.25$) derivatized with membranes containing the ganglioside G_{T1B} (89.5% DMOPC, 9% DMPs, 0.5% G_{T1B} , 1% Texas red–DPPE), which have been incubated with various concentrations of TT. Binding of TT to membrane surface G_{T1B} induces a condensed-to-dispersed phase transition, as detected in the $g(r)$ plots as well as by direct observation of the colloid (inset images). **c**, Series of experiments as in **b**, except with 0.5% G_{M1} in place of G_{T1B} . Binding of CTB to the G_{M1} membrane surface induces the transition. Incubation of CTB with G_{T1B} colloids or TT with G_{M1} colloids produced no effect.

The protein systems studied here include antibody binding membrane surface antigen, and two different bacterial toxin proteins, cholera (CTB) and tetanus (TT), binding their respective membrane ligands, monosialoganglioside G_{M1} and trisialoganglioside G_{T1B} . Beads coated with 9% phosphatidylserine membranes formed condensed phases, consisting of short-lived crystallites, for all membrane surface ligands studied. In all cases, protein binding to membrane surfaces triggered a condensed-to-dispersed phase transition. Figure 3 depicts a time sequence of a phase transition triggered by addition of protein at time $t = 0$ s. These experiments were performed with $\sim 300 \mu\text{l}$ of solution in ~ 5 -mm round wells of a 96-well plate. Within 30 s of adding a drop of protein solution to the top of the well, uniform disruption of the condensed phase was discernible. Within 60 s, the colloid attained a dispersed distribution. Individual bead mobility is unaffected by protein binding. Exposure to a particular protein of interest triggered a phase transition only when the appropriate cognate ligand was incorporated into the colloid membrane. The physical presence of protein bound to the membrane surface increases the closest approach position between two membranes¹⁵ and, correspondingly, reduces the cumulative strength of the van der Waals attraction between beads.

Quantitative analysis of the colloidal phases was performed by extracting the pair distribution function, $g(r)$. Bead positions were measured from wide-field ($\sim 1 \text{ mm}^2$) images by an object-locating algorithm to a precision of ~ 16 nm. Experiments were generally carried out at bead area fractions of $\phi \approx 0.15$ – 0.25 , corresponding to $\sim 7,000$ individual beads per image. 5–10 independent images were averaged to generate $g(r)$ from $>10^8$ measured pair distances. For a finite rectangular window of spatial dimensions X by Y , $g(r)$ can be computed from

$$g(r) = \frac{\eta(r)X^2Y^2}{N(N-1)\delta r[\pi XYr - 2(X+Y)r^2 + r^3]} \quad (1)$$

where $\eta(r)$ is the number of bead pairs with separation distance $r \pm \delta r/2$ ($\delta r = 40$ nm for the data presented here) and N is the total number of beads (see Supplementary Information).

Plots of $g(r)$ for different time points during a phase transition triggered by protein binding are illustrated in Fig. 3b. The condensed phase $g(r)$ is characterized by a large peak at the nearest-neighbour separation distance of one bead diameter (r_0), and secondary peaks occurring at $r = \sqrt{3}r_0$ and $2r_0$, corresponding to next-nearest neighbours in the hexagonal crystallites. Independent measurements of $g(r)$ were highly consistent. Standard deviations in the magnitude of the r_0 peak determined from separate colloidal preparations were generally less than 5%. Dispersed phases, consisting of random distributions and correspondingly flat $g(r)$ functions, are visibly distinguishable from condensed phases. Quantitative determination of $g(r)$ additionally distinguishes a range of intermediate distributions. These can be transient, such as the dispersing crystallites in Fig. 3a, $t = 30$ s. Intermediate degrees of order were also observed in near-equilibrium distributions, corresponding to differing amounts of protein binding on the membrane surface.

Measurements of near-equilibrium colloidal distributions over a range of protein and ligand concentrations were performed to explore the utility of the phase transition as a readout of protein binding on membrane surfaces. Antibody studies were performed using a monoclonal IgG antibody that binds the fluorescent membrane probe, Texas red-DPPE. Samples incubated with $20 \mu\text{g ml}^{-1}$ anti-Texas red antibody exhibit a clear transition from condensed to dispersed phases for ligand surface concentrations $\geq 10^{-4}$ monolayer (Fig. 4a). This corresponds to about ten ligand molecules on each membrane within the contact region where intermembrane separations are <10 nm ($5\text{-}\mu\text{m}$ beads). For bacterial toxin binding studies, the ganglioside ligands G_{T1B} or G_{M1} were incorporated into membranes at a constant 0.5%. Incubation with TT triggered formation of a dispersed phase in the G_{T1B} -containing

colloid, while no effect was produced in the G_{M1} colloid (Fig. 4b). Analogously, exposure to CTB triggered the transition to a dispersed phase in the G_{M1} colloid without producing any effect on the G_{T1B} colloid (Fig. 4c). Exposure to bungarotoxin ($1 \mu\text{M}$) produced no effect in either colloid. In a parallel set of experiments on planar supported membranes, the effective dissociation constants for TT- G_{T1B} and CTB- G_{M1} binding under these conditions were directly measured to be ~ 60 and ~ 41 nM, respectively (see Supplementary Information). The magnitude of the r_0 peak in the measured $g(r)$ traces the surface concentration of bound protein.

The membrane-derivatized colloidal system introduced here has potential applications to a broad set of problems involving chemical events on cell membrane surfaces. These range from mapping ligand interactions with numerous cell surface proteins to detection of membrane-targeting bacterial toxins (for example, botulism, cholera, anthrax, tetanus) and viruses. Membrane-derivatized beads can be combined in heterogeneous mixtures or with live cells (for example, rosetting), allowing the methodology outlined here to be applied to analysis of intermembrane receptor–ligand interactions. Implementation of the colloid assay is amenable to automated liquid handling and imaging systems. Detailed analysis of spatial distribution functions, such as the pair distribution studied here, enables characterization of subtle interactions, including those that may not produce qualitatively recognizable effects. At the same time, the use of membrane coatings on colloidal particles offers an extensive repertoire of chemical functionality, which may prove valuable in non-biological settings. We anticipate that the general principles illustrated with lipid membranes in this work could be extended to other materials, such as the recently developed polymer vesicles¹⁶. □

Received 24 July; accepted 14 November 2003; doi:10.1038/nature02209.

1. Sackmann, E. Supported membranes: Scientific and practical applications. *Science* **271**, 43–48 (1996).
2. Groves, J. T. Membrane array technology for drug discovery. *Curr. Opin. Drug Discov. Dev.* **5**, 606–612 (2002).
3. Grakoui, A. et al. The immunological synapse: A molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).
4. Fang, Y., Frutos, A. G. & Lahiri, J. Membrane protein microarrays. *J. Am. Chem. Soc.* **124**, 2394–2395 (2002).
5. Hoffman, T. L., Canziani, G., Jia, L., Rucker, J. & Doms, R. W. A biosensor assay for studying ligand-membrane receptor interactions: binding of antibodies and HIV-1 Env to chemokine receptors. *Proc. Natl Acad. Sci. USA* **97**, 11215–11220 (2000).
6. Yang, T., Jung, S.-y., Mao, H. & Cremer, P. S. Fabrication of phospholipid bilayer-coated microchannels for on-chip immunoassays. *Anal. Chem.* **73**, 165–169 (2001).
7. Anderson, V. J. & Lekkerkerker, H. N. W. Insights into phase transition kinetics from colloid science. *Nature* **416**, 811–815 (2002).
8. Yethiraj, A. & van Blaaderen, A. A colloidal model system with an interaction tunable from hard sphere to soft and dipolar. *Nature* **421**, 513–517 (2003).
9. Ramos, L., Lubensky, T. C., Dan, N., Nelson, P. & Weitz, D. A. Surfactant-mediated two-dimensional crystallization of colloidal crystals. *Science* **286**, 2325–2328 (1999).
10. Dinsmore, A. D. et al. Colloidosomes: selectively permeable capsules composed of colloidal particles. *Science* **298**, 1006–1009 (2002).
11. Manoharan, V. N., Elsesser, M. T. & Pine, D. J. Dense packing and symmetry in small clusters of microspheres. *Science* **301**, 483–487 (2003).
12. Yang, P. et al. Hierarchically ordered oxides. *Science* **282**, 2244–2246 (1998).
13. Bayerl, T. M. & Bloom, M. Physical properties of single phospholipid bilayers adsorbed to micro glass beads. *Biophys. J.* **58**, 357–362 (1990).
14. Buranda, T. et al. Biomimetic molecular assemblies on glass and mesoporous silica microbeads for biotechnology. *Langmuir* **19**, 1654–1663 (2003).
15. Wong, A. P. & Groves, J. T. Molecular topography imaging by intermembrane fluorescence resonance energy transfer. *Proc. Natl Acad. Sci. USA* **99**, 14147–14152 (2002).
16. Discher, D. E. & Eisenberg, A. Polymer vesicles. *Science* **297**, 967–973 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Parthasarathy and N. Clack for discussions, and D. Discher for advice on the manuscript. M.J. was supported by a Fulbright scholarship and by the Czech Republic Ministry of Education. This work was supported in part by the Burroughs Wellcome Career Award in the Biomedical Sciences (to J.T.G.) and by the US Department of Energy.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.T.G. (JTGroves@lbl.gov).

Critically pressured free-gas reservoirs below gas-hydrate provinces

Matthew J. Hornbach, Demian M. Saffer & W. Steven Holbrook

Department of Geology and Geophysics, University of Wyoming, Laramie, Wyoming 82071, USA

Palaeoceanographic data have been used to suggest that methane hydrates play a significant role in global climate change. The mechanism by which methane is released during periods of global warming is, however, poorly understood¹. In particular, the size and role of the free-gas zone below gas-hydrate provinces remain relatively unconstrained, largely because the base of the free-gas zone is not a phase boundary and has thus defied systematic description. Here we evaluate the possibility that the maximum thickness of an interconnected free-gas zone is mechanically regulated by valving caused by fault slip in overlying sediments². Our results suggest that a critical gas column exists below most hydrate provinces in basin settings, implying that these provinces are poised for mechanical failure and are therefore highly sensitive to changes in ambient conditions³. We estimate that the global free-gas reservoir may contain from one-sixth to two-thirds of the total methane trapped in hydrate⁴. If gas accumulations are critically thick along passive continental slopes, we calculate that a 5 °C temperature increase at the sea floor could result in a release of ~2,000 Gt of methane from the free-gas zone, offering a mechanism for rapid methane release during global warming events.

Gas hydrate, a solid consisting of water and mostly methane gas that is stable at high pressures and low temperatures, exists to a maximum depth of a few hundred metres below the sea floor of every continental margin on Earth. Because temperature increases with depth, gas hydrate is stable only in the upper few hundred metres of continental margin sediments, below which gas and water are stable. A bottom-simulating reflection (BSR)—a strong, negative-polarity seismic reflection that mimics the sea floor in seismic reflection profiles—indicates the base of hydrate stability and the top of the underlying free-gas zone (FGZ; Fig. 1). Because as much as 10,000 Gt of carbon may be trapped in marine methane hydrate, hydrate dissociation and free-gas migration into the biosphere may play a role in global climate^{5,6}, but the nature and mechanisms of methane escape are poorly understood. In particular, the size of the free-gas reservoir and its role in gas escape and climate change are unclear¹. As a result, climate models lack a quantifiable mechanism for including the role of free gas in climate change events¹.

Recent studies^{7,8} demonstrate that free-gas migration may occur at hydrate provinces via gas chimneys and along faults. Here we calculate the amount of interconnected free gas below the BSR necessary to trigger fault slip in the overlying hydrate-bearing sediments, allowing gas to migrate out of the FGZ and potentially through the hydrate stability zone (Fig. 1). Such free-gas columns may exist within permeable flat-lying sediments or dipping layers, or along faults^{7–9}. We first evaluate this scenario for the Blake ridge, a sediment drift deposit located 300 km east of South Carolina, using constraints on FGZ thickness from sonic well logs. We then calculate the thickness of the FGZ required for fault reactivation in ocean basins, as a function of BSR depth and sediment properties, and compare our results with measured gas column heights below BSRs worldwide. Though this model does not predict FGZ thickness for hydrate provinces in tectonic settings where the maximum principal stress (σ_1) is near horizon-

tal, we compare measured FGZ thickness below passive basin regions and active tectonic regions, and note a systematic difference in free-gas thickness below the two populations. We use this information to estimate an upper bound on the global inventory of free gas below hydrate provinces.

To calculate the gas column height required for fault slip in passive-margin marine basin settings, we assume that $\sigma_1 = \sigma_v$ (at or near vertical), and the minimum principal stress, $\sigma_3 = \sigma_h$ (at or near horizontal), as indicated by low sea-floor slopes (1–2°) and conjugate faults dipping ~60° found throughout passive basin settings⁷. We calculate σ_v (total vertical stress) from bulk density data collected by drilling, and use a poroelastic formulation to estimate total horizontal stress, σ_h , which accounts for horizontal stresses due to both the Poisson's effect and pore pressure increase in a laterally confined basin¹⁰:

$$\sigma_v = \int_0^z \rho_b(z)gzdz \quad (1)$$

$$\sigma_h = [\nu/(1 - \nu)]\sigma_v + \alpha[(1 - 2\nu)/(1 - \nu)](P_w + P_{\text{gas}}) \quad (2)$$

where z is thickness of sediment above the BSR, $\rho_b(z)$ is the sediment bulk density, g is gravitational acceleration, ν is the Poisson's ratio of sediments at the BSR, α is the Biot coefficient of effective stress, and P_w and P_{gas} are pressures of water and gas, respectively. The critical gas pressure (P_{crit}) necessary to initiate fault slip is given by:

$$P_{\text{crit}} = [(\sigma_v + \sigma_h)/2] + [(\sigma_v - \sigma_h)/2](1 + \cos 2\theta - \sin 2\theta/\mu) + C/\mu - P_w \quad (3)$$

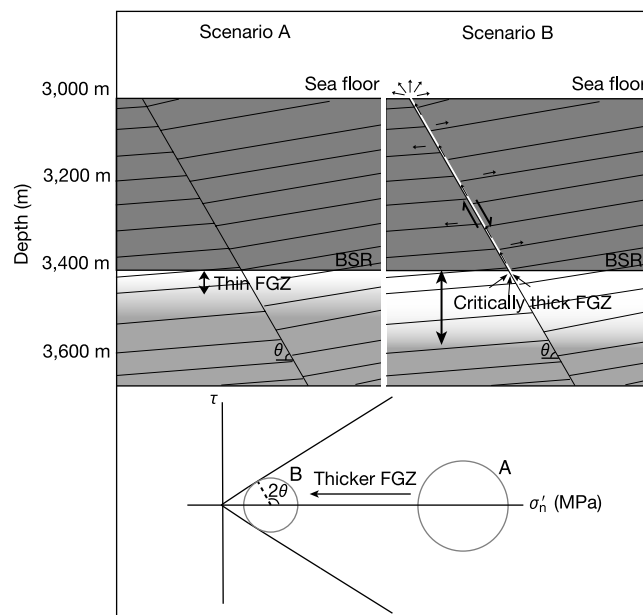


Figure 1 Cross-section cartoon of a normal fault in a basin. Scenario A, a thin FGZ below the BSR (top left panel); scenario B, a thick FGZ below the BSR (top right panel) where gas overpressures at the BSR interface can cause fault reactivation and gas escape. The arrows on the plot of scenario B indicate possible gas migration pathways during or immediately following fault slip. Bottom, Mohr diagram for sediments immediately below the BSR, demonstrating the two scenarios. For shallow submarine sediments, cohesion is near zero. Increased gas column thickness produces higher gas pressure at the BSR, reducing effective normal stress (σ'_n) and moving the Mohr's circle to the left. Note that circle diameter decreases as it moves left owing to poroelasticity.

where θ is the fault dip angle, C is the cohesive strength (assumed equal to zero), and μ is the coefficient of sliding friction. We determined $\theta = 59^\circ$ for the Blake ridge by measuring the dips of 150 normal faults from seismic data, and calculated μ by $\mu = \tan(90 - 2\theta) = 0.53$. For global calculations of P_{crit} , we assume that P_w is hydrostatic and allow ν to vary from 0.41 to 0.45, in accordance with measured values for sediments at the Blake ridge based on drilling results and with global compilations for marine sediments^{11–14}. We allow α to vary from 0.67 to 0.77, corresponding to 25–0% bulk hydrate in sediment, respectively¹¹. Values of σ_h calculated using these parameters, and assuming no gas- or water-phase overpressures, correspond to a σ_h/σ_v ratio of ~ 0.75 . To calculate the FGZ height necessary to trigger brittle failure, we assume that all gas is methane, and account for failure of the ideal gas law at high pressures by calculating the compressibility factor necessary for non-ideal compression of methane gas¹⁵. Values of critical FGZ height calculated by equation (3) reflect an upper-bound on interconnected FGZ thickness. It is important to note that other factors, including gas availability and solubility, could result in subcritical FGZ thickness, whereas significant capillary entry pressure or fluid water-phase overpressure could reduce the critical gas column thickness required for failure³.

Our results reveal that a critical gas column thickness exists below the Blake ridge, suggesting that faults at the crest of the ridge are poised at the mechanical limit of failure (Fig. 2). Fault reactivation should occur when gas column thickness below the BSR exceeds 150–290 m. Vertical seismic profiles (VSPs) and well logs from two drill sites on the Blake ridge, ODP sites 995 and 997, show that gas creates a low-velocity zone, 200–250 m thick, below the BSR⁹. As long as the gas below the BSR forms an interconnected network, faults that intersect the BSR on the Blake ridge are poised for reactivation. Changes in external conditions, such as pressure and temperature, therefore have the potential to release methane gas from beneath the BSR into the hydrate stability zone, and perhaps into the oceans and atmosphere³.

To test the hypothesis that FGZ thickness is mechanically regu-

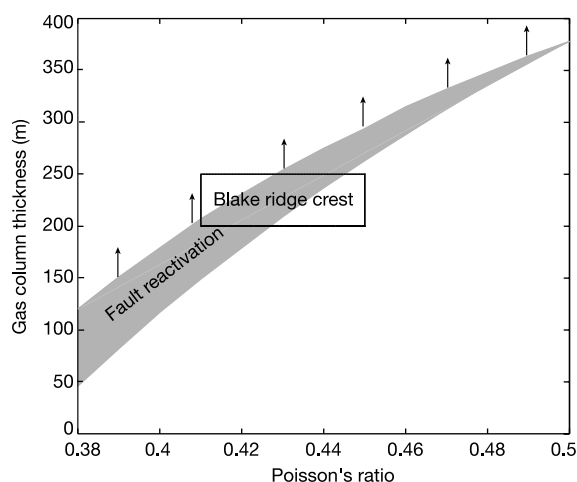


Figure 2 Gas column thickness required for fault reactivation (grey) at the Blake ridge. Values are calculated for a water depth of 3,000 m and BSR depth of 400 m—the values for the location of the ODP Leg 164 drill sites. The top and bottom boundaries of the grey area correspond to values of $\alpha = 0.77$ (for 0% hydrate) and of $\alpha = 0.67$ (for 25% bulk hydrate in sediment). Any increase in hydrate content above 25% would cause the envelope to shift downward. Arrows indicate the amount that failure thresholds will shift for 100 kPa of fault cohesion at the BSR, which might result from hydrate cementation. Box indicates range of Poisson ratios and gas column heights measured on the Blake ridge^{9,13}.

lated globally, we calculate critical FGZ thickness as a function of BSR depth and compare our results to nine hydrate provinces where conjugate normal faults indicate a near-vertical σ_1 , and where the FGZ thickness is well constrained by seismic analysis or drilling (Fig. 3). Of these, seven have FGZ thickness at or near the critical value, existing over a range of BSR depths from <200 to 650 m. Critical gas column thickness increases as a function of BSR (and hence water) depth, because σ_v and σ_h are correspondingly larger.

Observations are consistent with a mechanically regulated upper limit on FGZ thickness; importantly, no observed gas column thickness significantly exceeds the calculated critical value (Fig. 3). These observations also support the assumption that gas below hydrate basins forms an interconnected network, because vertically isolated gas pockets could occupy a zone much thicker than our predicted upper limit. We conclude for hydrates in passive basin settings that (1) the observed increase in gas column thickness as a function of BSR depth is consistent with a mechanically regulated FGZ; (2) although available data are limited, critical gas columns

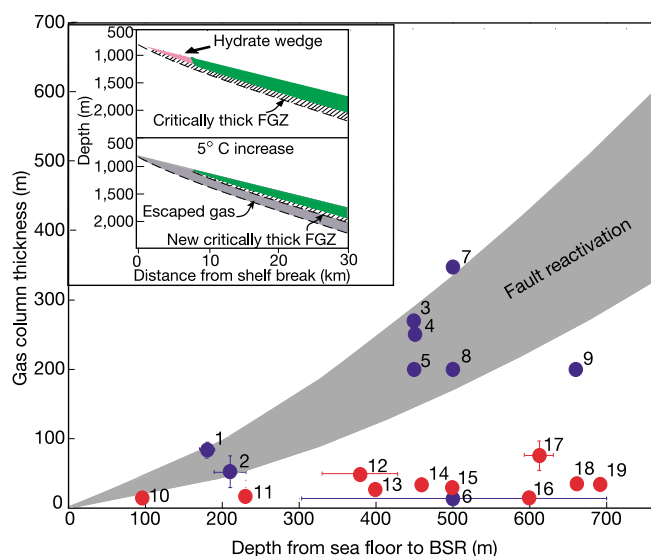


Figure 3 Depth from sea floor to BSR (m) versus gas column thickness. Blue symbols (numbered 1–9) show gas column thickness below BSRs at several hydrate provinces in basin settings where σ_1 is near vertical. Red symbols (10–19) show gas column thickness below BSRs at hydrate provinces in compressional settings where σ_1 is probably near horizontal. Note that gas column thickness generally increases with depth for the blue points, but remains nearly constant for the red points. The grey area indicates the interconnected FGZ thickness required for fault reactivation. The upper bound of the grey area corresponds to values of $\alpha = 0.77$ (zero hydrate at the BSR) and $\nu = 0.45$; the lower bound corresponds to low end values of $\alpha = 0.67$ (25% bulk hydrate) and $\nu = 0.41$. Point locations: 1, Barents Sea²⁵; 2, the Norwegian margin²⁶; 3, Bering Sea basin²⁷; 4, 5, Blake ridge⁹; 6, Beaufort Sea²⁸; 7–9, basin in Makran accretionary prism that has experienced rapid sedimentation^{29,30}; 10, offshore Cascadia¹⁷; 11, offshore Chile¹⁶; 12, Niger delta¹⁹; 13, 15, offshore Colombia¹⁸; 14, Gulf of Oman anticline²⁰; 16, offshore Peru²¹; 17, South Shetland islands²²; 18, 19, Oman accretionary toe^{31,32}. Top inset, model of the hydrate stability zone and critically thick FGZ below for a bottom-water temperature of $\sim 0^\circ\text{C}$. The BSR is located at the interface between the sediment overburden where hydrate can exist (green) and the FGZ (black and white stripes). The pink area marks the hydrate wedge. Bottom inset, new BSR depth and critical FGZ thickness for a 5°C temperature rise at the sea floor. The grey area denotes the area over which gas (and dissociated hydrate) must escape. Note the thinner critical FGZ due to bottom-water warming, which results in less sediment overburden above the BSR.

required for mechanical failure appear common; and (3) a reservoir of potentially mobile methane that is sensitive to changes in environmental conditions exists below the BSR.

A systematic difference in gas column thickness exists between hydrate provinces in basin settings where σ_1 is near vertical (Fig. 3, blue points), and compressional settings where σ_1 is probably near horizontal (Fig. 3, red points). Unlike FGZs in basin hydrate provinces—which increase in thickness with BSR depth—FGZs in compressional settings are generally thinner and more constant in thickness, averaging ~35 m thick, and typically not exceeding 50 m regardless of BSR depth (Fig. 3, red points only)^{16–22}. One possible explanation for thin FGZs in active tectonic settings is that they are often characterized by water phase overpressures, which reduce the gas column height required to initiate slip³.

The concept of a mechanical upper limit on FGZ thickness allows calculation of the global potential volume of free methane gas beneath hydrate provinces. Following Dickens' model¹ for global hydrate volume calculation, we calculate a mechanically regulated maximum FGZ thickness along a typical continental margin (Fig. 3 top inset). To estimate an upper limit on volume of the global free-gas reservoir below marine hydrate provinces we assume (1) that passive margins comprise 50% of all margins, (2) 30% of all passive margins have critical gas columns⁴, (3) sediment porosity averages 50%, (4) gas fills 4% of the underlying pore space^{19,21}, and (5) gas column height on active margins averages 35 m, as indicated by existing data. These assumptions yield a maximum total free-gas inventory of ~1,550 Gt, or about one-sixth to two-thirds of the total methane trapped in hydrate, depending on which global estimate of hydrate volume is used^{4,23}. A reasonable lower limit of the global free-gas inventory is ~150 Gt, assuming that only 10% of all passive margins have critical gas columns, and that free gas fills only 1% of the pore space^{9,16}. An intermediate estimate of the mass of the global free-gas reservoir in hydrate provinces, calculated assuming critical gas columns on 20% of passive margins and assuming 2% free-gas saturation, is ~520 Gt.

Our model offers a simple physical mechanism to explain methane gas escape during bottom-water warming events, such as the Late Palaeocene Thermal Maximum (LPTM). Dickens¹ demonstrates that a 5°C bottom-water temperature increase may result in the dissociation of ~2,000 Gt of methane from hydrate—the amount necessary to explain the observed $\delta^{13}\text{C}$ excursion (–3‰) during the LPTM. This model, however, assumes that the majority of the methane that escapes during the warming event comes from dissociation of the hydrate wedge—a narrow area on the continental slope where hydrates from the sea floor to the BSR would dissociate completely during the warming event (Fig. 3 inset). For the 'hydrate wedge' to supply the methane necessary to drive the $\delta^{13}\text{C}$ excursion, hydrate would have to fill, on average, $\geq 15\%$ of the pore space—a value considered unreasonably high^{1,9}. This suggests that significant quantities of dissociated free gas trapped below hydrate provinces must also escape to cause the LPTM $\delta^{13}\text{C}$ excursion¹.

Using Dickens' model¹ of BSR depths for bottom-water temperatures of 0° and 5°C, we calculate the amount of free gas produced by hydrate dissociation that must escape the FGZ to keep gas pressures from exceeding critical values. The minimum amount of methane gas that must escape, calculated using our minimum free-gas estimate, and assuming that hydrate fills only 2% of the sediment pore space above the BSR⁹, is 250 Gt. A probable upper limit on gas escape, calculated using our maximum free-gas estimate and assuming that hydrate fills 10% of the sediment pore space²¹, is 3,800 Gt. The large difference between these estimates is due to the concentration of methane in hydrate. Therefore, ~2,000 Gt represents a plausible estimate for free-gas escape from the hydrate wedge and FGZ into the hydrate stability zone—and perhaps the oceans and atmosphere—for a 5°C bottom-water temperature increase (Fig. 3 inset). This estimate is conservative, in that it does not include the possibility that gas overpressures

could trigger submarine landslides, which could prompt complete gas evacuation^{3,24}. □

Received 16 June; accepted 5 November 2003; doi:10.1038/nature02172.

- Dickens, G. R. The potential volume of oceanic methane hydrates with variable external conditions. *Org. Geochem.* **32**, 1179–1193 (2001).
- Finkbeiner, T., Zoback, M., Flemings, P. & Stump, B. Stress, pore pressure, and dynamically constrained hydrocarbon columns in the South Eugene Island field, northern Gulf of Mexico. *Bull. Am. Assoc. Petrol. Geol.* **85**, 1007–1031 (2001).
- Flemings, P., Liu, X. & Winters, W. J. Critical pressure and multiphase flow in Blake Ridge gas hydrates. *Geology* **31**, 1057–1060 (2003).
- Milkov, A. V. Global estimates of hydrate-bound gas in marine sediments: How much is really out there? *Earth Sci. Rev.* (in the press).
- Dickens, G. R., O'Neill, J. R., Rea, D. K. & Owen, R. M. Dissociation of oceanic methane hydrate as a cause of the carbon isotope excursion at the end of the Paleocene. *Paleoceanography* **10**, 965–971 (1995).
- Katz, M. E., Pak, D. K., Dickens, G. R. & Miller, K. G. The source and fate of massive carbon input during the latest Paleocene Thermal Maximum. *Science* **286**, 1531–1533 (1999).
- Gorman, A. R. *et al.* Migration of methane gas through the hydrate stability zone in a low-flux hydrate province. *Geology* **30**, 327–330 (2002).
- Tryon, M. D., Brown, K. M. & Torres, M. E. Fluid and chemical flux in and out of sediments hosting methane hydrate deposits on Hydrate Ridge, OR, II: Hydrological processes. *Earth Planet. Sci. Lett.* **201**, 541–557 (2002).
- Holbrook, W. S., Hoskins, H., Wood, W. T., Stephen, R. A. & Lizzarralde, D. Methane gas-hydrate and free gas on the Blake Ridge from vertical seismic profiling. *Science* **273**, 1840–1843 (1996).
- Engelder, T. & Fischer, M. P. Influence of poroelastic behavior on the magnitude of minimum horizontal stress, S_{h} , in overpressured parts of sedimentary basins. *Geology* **22**, 949–952 (1994).
- Helgerud, M. B., Dvorkin, J. & Nur, A. Elastic-wave velocity in marine sediments with gas hydrates: Effective medium modeling. *Geophys. Res. Lett.* **26**, 2021–2024 (1999).
- Miller, A. T. (ed.) *Proc ODP Init. Rep. Leg 164*, 179–198, (1996).
- Guerin, G., Goldberg, D. & Meltzer, A. Characterization of in situ elastic properties of gas hydrate-bearing sediments on the Blake Ridge. *J. Geophys. Res.* **104**, 17781–17795 (1999).
- Hamilton, E. L. V_p/V_s and Poisson's ratios in marine sediments and rocks. *J. Acoust. Soc. Am.* **66**, 1093–1101 (1979).
- Mattar, L. & Brar, G. S. Compressibility of natural gases. *J. Can. Petrol. Technol.* **14**, 77–80 (1975).
- Bangs, N. L. B., Sawyer, D. S. & Golovchenko, X. Free gas at the base of the gas hydrate zone in the vicinity of the Chile triple junction. *Geology* **21**, 905–908 (1993).
- Hyndman, R. D., Spence, G. D., Chapman, R., Riedel, M. & Edwards, R. N. in *Natural Gas Hydrates: Occurrence, Distribution, and Detection* (eds Paul, C. K. & Dillon, W. P.) 273–295 (American Geophysical Union, Washington DC, 2001).
- Minshull, T. A., Singh, S. C. & Westbrook, G. K. Seismic velocity structure at a gas hydrate reflector, offshore western Columbia, from full waveform inversion. *J. Geophys. Res.* **99**, 4715–4734 (1994).
- Hovland, M., Gallagher, J. W., Clennell, M. B. & Lekvam, K. Gas hydrate and free gas volumes in marine sediments: Example from the Niger Delta front. *Mar. Petrol. Geol.* **14**, 245–255 (1997).
- Minshull, T. A. & White, R. Sediment compaction and fluid migration in the Makran accretionary prism. *J. Geophys. Res.* **94**, 7387–7402 (1989).
- Pecher, I. A., Minshull, T. A., Singh, S. C. & von Huene, R. Velocity structure of a bottom simulating reflector offshore Peru: Results from full waveform inversion. *Earth Planet. Sci. Lett.* **139**, 459–469 (1996).
- Tinivella, U., Lodolo, E., Camerlenghi, A. & Boehm, G. in *Gas Hydrates: Relevance to World Margin Stability and Climate Change* (eds Henriot, J. P. & Mienert, J.) 141–151 (The Geological Society, London, 1998).
- Kvenvolden, K. A. Gas hydrates—geological perspective and global change. *Rev. Geophys.* **31**, 173–187 (1993).
- Paull, C. K., Buelow, W. J., Bussler, W. III & Borowski, W. S. Increased continental-margin slumping frequency during sea-level lowstands above gas hydrate-bearing sediments. *Geology* **24**, 143–146 (1996).
- Andreassen, K., Hogstad, K. & Berteussen, K. A. Gas hydrate in the southern Barents Sea, indicated by a shallow seismic anomaly. *First Break* **8**, 237–245 (1990).
- Mienert, J., Andreassen, K., Posewang, J. & Lukas, D. in *Gas Hydrates: Challenges for the Future* (eds Holder, G. D. & Bishnoi, P. R.) 200–210 (New York Academy of Sciences, New York, 2000).
- Scholl, D. W. & Hart, P. E. The future of energy gases. *Prof. Pap. US Geol. Surv.* **1570**, 331–351 (1993).
- Andreassen, K., Hart, P. E. & Grantz, A. Seismic studies of a bottom simulating reflection related to gas hydrate beneath the continental margin of the Beaufort Sea. *J. Geophys. Res.* **100**, 12659–12673 (1995).
- Sain, K., Minshull, T. A., Singh, S. C. & Hobbs, R. W. Evidence for a thick free gas layer beneath the bottom simulating reflector in the Makran accretionary prism. *Mar. Geol.* **164**, 3–12 (2000).
- Grevenmeyer, I., Rosenberger, A. & Villinger, H. Natural gas hydrates on the continental slope off Pakistan: constraints from seismic techniques. *Geophys. J. Int.* **140**, 295–310 (2000).
- White, R. S. Gas hydrate layers trapping free gas in the Gulf of Oman. *Earth Planet. Sci. Lett.* **42**, 114–120 (1979).
- White, R. S. Seismic bright spots in the Gulf of Oman. *Earth Planet. Sci. Lett.* **37**, 29–37 (1977).

Acknowledgements We thank J. Dickens, P. Flemings and A. Milkov for discussions. This work was funded jointly by the US National Science Foundation and the US Department of Energy.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.H. (mhornbac@uwyo.edu).

Extinction risk from climate change

Chris D. Thomas¹, Alison Cameron¹, Rhys E. Green², Michel Bakkenes³, Linda J. Beaumont⁴, Yvonne C. Collingham⁵, Barend F. N. Erasmus⁶, Marínez Ferreira de Siqueira⁷, Alan Grainger⁸, Lee Hannah⁹, Lesley Hughes⁴, Brian Huntley⁵, Albert S. van Jaarsveld¹⁰, Guy F. Midgley¹¹, Lera Miles^{8*}, Miguel A. Ortega-Huerta¹², A. Townsend Peterson¹³, Oliver L. Phillips⁸ & Stephen E. Williams¹⁴

¹Centre for Biodiversity and Conservation, School of Biology, University of Leeds, Leeds LS2 9JT, UK

²Royal Society for the Protection of Birds, The Lodge, Sandy, Bedfordshire SG19 2DL, UK, and Conservation Biology Group, Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

³National Institute of Public Health and Environment, P.O. Box 1, 3720 BA Bilthoven, The Netherlands

⁴Department of Biological Sciences, Macquarie University, North Ryde, 2109, NSW, Australia

⁵University of Durham, School of Biological and Biomedical Sciences, South Road, Durham DH1 3LE, UK

⁶Animal, Plant and Environmental Sciences, University of the Witwatersrand, Private Bag 3, WITS 2050, South Africa

⁷Centro de Referência em Informação Ambiental, Av. Romeu Tórtima 228, Barão Geraldo, CEP:13083-885, Campinas, SP, Brazil

⁸School of Geography, University of Leeds, Leeds LS2 9JT, UK

⁹Center for Applied Biodiversity Science, Conservation International, 1919 M Street NW, Washington, DC 20036, USA

¹⁰Department of Zoology, University of Stellenbosch, Private Bag X1, Stellenbosch 7602, South Africa

¹¹Climate Change Research Group, Kirstenbosch Research Centre, National Botanical Institute, Private Bag x7, Claremont 7735, Cape Town, South Africa

¹²Unidad Occidente, Instituto de Biología, Universidad Nacional Autónoma de México, México, D.F. 04510 México

¹³Natural History Museum and Biodiversity Research Center, University of Kansas, Lawrence, Kansas 66045 USA

¹⁴Cooperative Research Centre for Tropical Rainforest Ecology, School of Tropical Biology, James Cook University, Townsville, QLD 4811, Australia

* Present address: UNEP World Conservation Monitoring Centre, 219 Huntingdon Road, Cambridge CB3 0DL, UK

Climate change over the past ~30 years has produced numerous shifts in the distributions and abundances of species^{1,2} and has been implicated in one species-level extinction³. Using projections of species' distributions for future climate scenarios, we assess extinction risks for sample regions that cover some 20% of the Earth's terrestrial surface. Exploring three approaches in which the estimated probability of extinction shows a power-law relationship with geographical range size, we predict, on the basis of mid-range climate-warming scenarios for 2050, that 15–37% of species in our sample of regions and taxa will be 'committed to extinction'. When the average of the three methods and two dispersal scenarios is taken, minimal climate-warming scenarios produce lower projections of species committed to extinction (~18%) than mid-range (~24%) and maximum-change (~35%) scenarios. These estimates show the importance of rapid implementation of technologies to decrease greenhouse gas emissions and strategies for carbon sequestration.

The responsiveness of species to recent^{1–3} and past^{4,5} climate change raises the possibility that anthropogenic climate change could act as a major cause of extinctions in the near future, with the Earth set to become warmer than at any period in the past 1–40 Myr (ref. 6). Here we use projections of the future distributions of 1,103 animal and plant species to provide 'first-pass' estimates of extinction probabilities associated with climate change scenarios for 2050.

For each species we use the modelled association between current climates (such as temperature, precipitation and seasonality) and present-day distributions to estimate current distributional

areas^{7–12}. This 'climate envelope' represents the conditions under which populations of a species currently persist in the face of competitors and natural enemies. Future distributions are estimated by assuming that current envelopes are retained and can be projected for future climate scenarios^{7–12}. We assume that a species either has no limits to dispersal such that its future distribution becomes the entire area projected by the climate envelope model or that it is incapable of dispersal, in which case the new distribution is the overlap between current and future potential distributions (for example, species with little dispersal or that inhabit fragmented landscapes)¹¹. Reality for most species is likely to fall between these extremes.

We explore three methods to estimate extinction, based on the species–area relationship, which is a well-established empirical power-law relationship describing how the number of species relates to area ($S = cA^z$, where S is the number of species, A is area, and c and z are constants)¹³. This relationship predicts adequately the numbers of species that become extinct or threatened when the area available to them is reduced by habitat destruction^{14,15}. Extinctions arising from area reductions should apply regardless of whether the cause of distribution loss is habitat destruction or climatic unsuitability.

Because climate change can affect the distributional area of each species independently, classical community-level approaches need to be modified (see Methods). In method 1 we use changes in the summed distribution areas of all species. This is consistent with the traditional species–area approach: on average, the destruction of half of a habitat results in the loss of half of the distribution area summed across all species restricted to that habitat. However, this analysis tends to be weighted towards species with large distributional areas. To address this, in method 2 we use the average proportional loss of the distribution area of each species to estimate the fraction of species predicted to become extinct. This approach is faithful to the species–area relationship because halving the habitat area leads on average to the proportional loss of half the distribution of each species. Method 3 considers the extinction risk of each species in turn. In classical applications of the species–area approach, the fraction of species predicted to become extinct is equivalent to the mean probability of extinction per species. Thus, in method 3 we estimate the extinction risk of each species separately by substituting its area loss in the species–area relationship, before averaging across species (see Methods). Our conclusions are not dependent on which of these methods is used. We use $z = 0.25$ in the species–area relationship throughout, given its previous success in predicting proportions of threatened species^{14,15}, but our qualitative conclusions are not dependent on choice of z (Supplementary Information). As there are gaps in the data (not all dispersal/climate scenarios were available for each region), a logit–linear model is fitted to the extinction risk data to produce estimates for missing values in the extinction risk table (Table 1). Balanced estimates of extinction risk, averaged across all data sets, can then be calculated for each scenario.

For projections of maximum expected climate change, we estimate species-level extinction across species included in the study to be 21–32% (range of the three methods) with universal dispersal, and 38–52% for no dispersal (Table 1). For projections of mid-range climate change, estimates are 15–20% with dispersal and 26–37% without dispersal (Table 1). Estimates for minimum expected climate change are 9–13% extinction with dispersal and 22–31% without dispersal. Projected extinction varies between parts of the world and between taxonomic groups (Table 1), so our estimates are affected by the data available. The species–area methods differ from one another by up to 1.41-fold (method 1 versus method 3) in estimated extinction, whereas the two dispersal scenarios produce a 1.98-fold difference, and the three climate scenarios generate 2.05-fold variation.

Given its role in conservation planning, we also use a different approach to estimate extinction, modified from the IUCN Red Data Book listing procedure¹⁶: this is semi-numerical and includes components of expert judgement. Species are assigned to different threat categories based on distribution sizes and declines, with each category carrying a specified probability of extinction¹⁶ (see Methods and Supplementary Information). For scenarios of maximum expected climate change, 33% (with dispersal) and 58% (without dispersal) of species are expected to become extinct (Table 1). For mid-range climate change scenarios, 19% or 45% (with or without dispersal) of species are expected to become extinct, and for minimum expected climate change 11% or 34% (with or without dispersal) of species are projected to become extinct.

We can compare these values with the proportions of species projected to become extinct as the result of global habitat losses, currently the most widely recognized extinction threat. We apply the species–area relationship to changes in global land use that have taken place since human land conversion began¹⁷. Estimates of extinction range from 1% to 29%, depending on the biome (considering only species restricted to single biomes; Table 2). Given that a high proportion of the world's species reside in tropical forests (extinction estimate 4%; Table 2), global extinction related

to habitat loss would be expected to be in the lower half of the range, and thus lower than the rate projected for scenarios of mid-range climate change (24%; average of area methods). Projected conversion of humid tropical forest at an annual rate of 0.43% (ref. 18) from 1990 to 2050 predicts a further 6.3% of species committed to extinction.

Regional differences are expected, so we also compare the relative risks during 2000–2050 associated with land use and climate change (using area approaches) for the three region–taxon combinations that correspond most closely to single habitat or biome types. First, for montane Queensland forests¹², extinction risk is dominated by climate change (7–13% and 43–58% predicted extinction for minimum and maximum climate scenarios, respectively; 0% predicted on the basis of further habitat destruction, given its legal protection). Second, for cerrado vegetation in Brazil, high rates of habitat destruction¹⁹ make it possible that only current reserves will survive. Making this pessimistic assumption, an estimated additional 34% of all original species will be committed to extinction due to habitat destruction during 2000–2050, a value lower than the 48–56% of woody plant species projected to be committed to extinction for mid-range climate warming (38–45% for minimum warming). Last, for South African Proteaceae, 27% of all original species are projected to become extinct as a result of land use

Table 1 Projected percentage extinctions for different taxa and regions

Taxon	Region	With dispersal			No dispersal		
		Minimum expected climate change	Mid-range climate change	Maximum expected climate change	Minimum expected climate change	Mid-range climate change	Maximum expected climate change
Mammals	Mexico <i>n</i> = 96	2, 4, 5 5	2, 5, 7 8	–	9, 14, 18 24	10, 15, 20 26	–
	Queensland <i>n</i> = 11	10, 13, 15 16	–	48, 54, 80 77	–	–	–
	South Africa <i>n</i> = 5	–	24, 32, 46 0	–	–	28, 36, 59 69	–
Birds	Mexico <i>n</i> = 186	2, 2, 3 4	3, 3, 4 5	–	5, 7, 8 9	5, 7, 8 8	–
	Europe <i>n</i> = 34	–	–	4, 6, 6 7	–	–	13, 25, 38 48
	Queensland <i>n</i> = 13	7, 9, 10 12	–	49, 54, 72 85	–	–	–
	South Africa <i>n</i> = 5	–	28, 29, 32 0	–	–	33, 35, 40 51	–
Frogs	Queensland <i>n</i> = 23	8, 12, 18 13	–	38, 47, 67 68	–	–	–
Reptiles	Queensland <i>n</i> = 18	7, 11, 14 9	–	43, 49, 64 76	–	–	–
	South Africa <i>n</i> = 26	–	21, 22, 27 0	–	–	33, 36, 45 59	–
Butterflies	Mexico <i>n</i> = 41	1, 3, 4 7	3, 4, 5 7	–	6, 9, 11 13	9, 12, 15 19	–
	South Africa <i>n</i> = 4	–	13, 7, 8 0	–	–	35, 45, 70 78	–
	Australia <i>n</i> = 24	5, 7, 7 7	13, 15, 16 23	21, 22, 26 33	9, 11, 12 16	18, 21, 23 35	29, 32, 36 54
Other invertebrates	South Africa <i>n</i> = 10	–	18, 15, 24 0	–	–	28, 46, 80 85	–
Plants	Amazonia <i>n</i> = 9	–	–	44, 36, 79 69	–	–	100, 100, 99 87
	Europe <i>n</i> = 192	3, 4, 5 6	3, 5, 6 7	4, 5, 6 8	9, 11, 14 18	10, 13, 16 22	13, 17, 21 29
	Cerrado <i>n</i> = 163	–	–	–	38, 39, 45 66	48, 48, 57 75	–
	South Africa Proteaceae <i>n</i> = 243	–	24, 21, 27 38	–	–	32, 30, 40 52	–
All species		9, 10, 13 11 <i>n</i> = 604	15, 15, 20 19 <i>n</i> = 832	21, 23, 32 33 <i>n</i> = 324	22, 25, 31 34 <i>n</i> = 702	26, 29, 37 45 <i>n</i> = 995	38, 42, 62 58 <i>n</i> = 259

Projected percentage extinction values are given, based on species–area (for $z = 0.25$) and Red Data Book (bold) approaches. The three species–area estimates are ordered in each cell with method 1 given first, followed by method 2, then method 3. Values for 'All species' are based on both these raw values and estimates interpolated for the empty (–) cells (see Methods). In each instance, *n* is the number of species assessed directly.

Table 2 Estimated eventual extinction based on habitat loss

Biome	Percentage of world surface area (from ref. 17)			Percentage of species expected to go extinct by the species-area approach ($z = 0.25$)
	Undisturbed	1990	Area lost	
Cropland	0.0	10.9	0.0	0.0
Pasture	0.0	23.1	0.0	0.0
Ice	1.7	1.7	0.0	0.0
Tundra	4.8	4.6	0.2	1.0
Wooded tundra	2.0	1.9	0.1	1.1
Boreal forest	13.0	12.5	0.5	0.9
Cool conifer forest	2.7	2.1	0.6	6.1
Temperate mixed forest	5.2	2.2	3.0	19.2
Temperate deciduous forest	4.5	1.5	3.0	24.2
Warm mixed forest	4.7	1.9	2.8	20.3
Grassland/steppe	13.7	6.9	6.8	15.7
Hot desert	14.9	11.8	3.1	5.6
Scrubland	7.3	1.9	5.4	28.9
Savannah	11.9	6.2	5.7	15.1
Tropical woodland	6.1	4.4	1.7	8.0
Tropical forest	7.6	6.4	1.1	4.0

changes during 2000–2050 (for a pessimistic linear extrapolation of land use scenarios after 2020)²⁰, falling between the 30–40% (without dispersal) and 21–27% (with ubiquitous dispersal, which is unlikely for these plants) projected extinction for mid-range climate scenarios.

Many unknowns remain in projecting extinctions, and the values provided here should not be taken as precise predictions. Analyses need to be repeated for larger samples of regions and taxa, and the selection of climate change scenarios need to be standardized. Some of the most important uncertainties follow (see also Supplementary Information). We estimate proportions of species committed to future extinction as a consequence of climate change over the next 50 years, not the number of species that will become extinct during this period. Information is not currently available on time lags between climate change and species-level extinctions, but decades might elapse between area reduction (from habitat loss) and extinction¹⁴. Land use should also be incorporated into analyses: extinction risks might be higher than we project if future locations of suitable climate do not coincide with other essential resources (such as soil type or food resources). There is also uncertainty over which species will inhabit parts of the world projected to have climates for which no current analogue exists⁶. Equally importantly, all parts of the world will have historically unprecedented CO₂ levels⁶, which will affect plant species and ecosystems^{21,22} and herbivores²³, resulting in novel species assemblages and interactions.

Despite these uncertainties, we believe that the consistent overall conclusions across analyses establish that anthropogenic climate warming at least ranks alongside other recognized threats to global biodiversity. Contrary to previous projections²⁴, it is likely to be the greatest threat in many if not most regions. Furthermore, many of the most severe impacts of climate-change are likely to stem from interactions between threats, factors not taken into account in our calculations, rather than from climate acting in isolation. The ability of species to reach new climatically suitable areas will be hampered by habitat loss and fragmentation, and their ability to persist in appropriate climates is likely to be affected by new invasive species.

Minimum expected (that is, inevitable) climate-change scenarios for 2050 produce fewer projected 'committed extinctions' (18%; average of the three area methods and the two dispersal scenarios) than mid-range projections (24%), and about half of those predicted under maximum expected climate change (35%). These scenarios would diverge even more by 2100. In other words, minimizing greenhouse gas emissions and sequestering carbon²⁵ to realize minimum, rather than mid-range or maximum, expected climate warm-

ing could save a substantial percentage of terrestrial species from extinction. Returning to near pre-industrial global temperatures as quickly as possible could prevent much of the projected, but slower-acting, climate-related extinction from being realized. □

Methods

Climate-envelope modelling

The statistical match between climate variables and the boundaries of a species' distribution (climate envelope) represents conditions in which a species (normally) shows a positive demographic balance (rarely the absolute physical limits of a species, but the set of conditions under which it survives in at least some multi-species communities). The statistical approach is generic, but specific methods vary between studies (Supplementary Information). The approach has been validated by successfully predicting distributions of invading species when they arrive in new continents and by predicting distributional changes in response to glacial climate changes; its scope has been discussed widely (see, for example, refs 12, 26–29). Dispersal is assumed to be universal or zero (main text), except for the Mexican study in which 'universal dispersal' is movement through contiguous habitats¹¹.

Climate scenarios

Climate projections for 2050 were divided into three categories: minimum expected change resulting in a mean increase in global temperature of 0.8–1.7 °C and in CO₂ of 500 p.p.m. by volume (p.p.m.v.); mid-range scenarios with temperature increases of 1.8–2.0 °C and CO₂ increases of 500–550 p.p.m.v.; and maximum expected scenarios with temperature increases of >2.0 °C and CO₂ increases >550 p.p.m.v. (ref. 30). Projections for the year 2100 were allocated to 2050 scenarios according to their end temperatures and CO₂ levels (Supplementary Information).

Species

Within each region we use only data for endemic species (near-endemic in two cases). Near-endemics are defined as >90% of the distribution area known to occur (European birds) or thought to occur (cerrado plants, given incomplete data) within the region modelled. For European birds, near-endemics are included only if their extra-European distribution is similar to climate space within Europe. The focus on endemics permits us to model all range boundaries of each species (Supplementary Information).

Species-area approaches

Method 1 analyses overall changes in distribution areas, summed across species. The proportion of species in a region going extinct (E_1) is estimated as

$$E_1 = 1 - (\Sigma A_{\text{new}} / \Sigma A_{\text{original}})^z$$

where A_{original} is the area initially occupied by a species, and A_{new} is the future area projected for the same species, with summation carried out across species.

Method 2 is based on the average proportional change in distribution area, averaged across species. Regional extinction risk (E_2) is

$$E_2 = 1 - \{(1/n) \{ \Sigma (A_{\text{new}} / A_{\text{original}}) \} \}^z$$

where n is the number of species and $A_{\text{new}} / A_{\text{original}}$ is the proportional distribution change for each species in turn.

Method 3 estimates the extinction risk of each species in turn, averaging across species to derive regional estimates of extinction (E_3):

$$E_3 = (1/n) \Sigma [1 - (A_{\text{new}} / A_{\text{original}})^z]$$

Species for which $A_{\text{new}} > A_{\text{original}}$ were analysed as though $A_{\text{new}} = A_{\text{original}}$; that is, zero extinction would be returned by each equation if every species was projected to

expand (Supplementary Information). It is important to recognize that further work is required to establish empirically how the absolute and proportional area losses of individual species (in other words, the type of data from climate envelope projections) are related to extinction risk. As yet, no agreed standard method exists for such calculations: assumptions and uncertainties inherent in the three methods will be considered in detail elsewhere.

Extinction probability estimates were not available for all scenarios in every region/taxon, so means of scenarios were calculated after using a least-squares analysis of variance model to impute missing values. Region/taxon mean probabilities of extinction for each scenario were logit-transformed and a three-way analysis of variance was fitted (region/taxon $\times$ climate scenario $\times$ dispersal scenario; weighted by $\sqrt{N_{\text{species}}}$ per region/taxon study). The fitted model was used to impute expected values of the probability of extinction for those region/taxon and scenario combinations for which direct estimates were not available. Scenario means were then calculated from the combined direct estimates and imputed values, using $\sqrt{N_{\text{species}}}$ for each region/taxon as weights.

Red Data Book criteria

Each species is assigned to a threat category¹⁶, or classified 'Not Threatened' (0% risk), depending on the projected decline in area over 50 or 100 years (Supplementary Information) and the final distribution area. Existing areas were considered, so we present only the extra extinction attributable to climate change. Logit-transformed three-way analysis of variance was used to estimate extinction risks for empty cells, as with the species-area approaches.

Extinct: species with a projected future area of zero (100% of species assumed to be committed to eventual extinction).

Critically endangered: projected future distribution area $< 10 \text{ km}^2$, or decline by $> 80\%$ in 50 years (species assigned a 75% chance of extinction¹⁶).

Endangered: projected area $10\text{--}500 \text{ km}^2$, or $50\text{--}80\%$ decline in 50 years (species assigned a 35% chance of extinction¹⁶).

Vulnerable: projected area $500\text{--}2,000 \text{ km}^2$, or $> 50\%$ decline in 100 years on the basis of linear extrapolation of 50-year projection (species assigned a 15% chance of extinction¹⁶).

Received 10 September; accepted 13 October 2003; doi:10.1038/nature02121.

1. Parmesan, C. & Yohe, G. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* **421**, 37–42 (2003).
2. Root, T. L. *et al.* Fingerprints of global warming on wild animals and plants. *Nature* **421**, 57–60 (2003).
3. Pounds, J. A., Fogden, M. L. P. & Campbell, J. H. Biological response to climate change on a tropical mountain. *Nature* **398**, 611–615 (1999).
4. Overpeck, J., Whitlock, C. & Huntley, B. in *Paleoclimate, Global Change and the Future* (eds Alvarson, K., Bradley, R. & Pedersen, T.) 81–103 (Springer, Berlin, 2002).
5. Benton, M. J. & Twitchett, R. J. How to kill (almost) all life: the end-Permian extinction event. *Trends Ecol. Evol.* **18**, 358–365 (2003).
6. Houghton, J. T. *et al.* *Climate change 2001: the Scientific Basis. Contributions of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2001).
7. Bakkenes, M., Alkemade, J. R. M., Ihle, E., Leemans, R. & Latour, J. B. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. *Global Change Biol.* **8**, 390–407 (2002).
8. Beaumont, L. J. & Hughes, L. Potential changes in the distributions of latitudinally restricted Australian butterfly species in response to climate change. *Global Change Biol.* **8**, 954–971 (2002).
9. Erasmus, B. F. N., van Jaarsveld, A. S., Chown, S. L., Kshatriya, M. & Wessels, K. Vulnerability of South African animal taxa to climate change. *Global Change Biol.* **8**, 679–693 (2002).
10. Midgley, G. F., Hannah, L., Rutherford, M. C. & Powrie, L. W. Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. *Global Ecol. Biogeogr.* **11**, 445–451 (2002).
11. Peterson, A. T. *et al.* Future projections for Mexican faunas under global climate change scenarios. *Nature* **416**, 626–629 (2002).
12. Williams, S. E., Bolitho, E. E. & Fox, S. Climate change in Australian tropical rainforests: an impending environmental catastrophe. *Proc. R. Soc. Lond. B* **270**, 1887–1892 (2003).
13. Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
14. Brooks, T. M., Pimm, S. L. & Oyugi, J. O. Time lag between deforestation and bird extinction in tropical forest fragments. *Conserv. Biol.* **13**, 1140–1150 (1999).
15. Brooks, T. M., Pimm, S. L. & Collar, N. J. Deforestation predicts the number of threatened birds in insular Southeast Asia. *Conserv. Biol.* **11**, 382–394 (1997).
16. *IUCN Red List Categories and Criteria*, version 3.1. (IUCN Species Survival Commission, Gland, Switzerland, 2001).
17. Gaston, K. J., Blackburn, T. M. & Goldewijk, K. K. Habitat conversion and global avian biodiversity loss. *Proc. R. Soc. Lond. B* **270**, 1293–1300 (2003).
18. Achard, F. *et al.* Determination of deforestation rates of the world's humid tropical forests. *Science* **297**, 999–1002 (2002).
19. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
20. Roget, M., Richardson, D. M., Cowling, R. M., Lloyd, J. W. & Lombard, A. T. Current patterns of habitat transformation and future threats to biodiversity in terrestrial ecosystems of the Cape Floristic Region, South Africa. *Biol. Conserv.* **112**, 63–85 (2003).
21. Woodward, F. I. Potential impacts of global elevated CO_2 concentrations on plants. *Curr. Opin. Plant Biol.* **5**, 207–211 (2002).
22. Bond, W. J., Midgley, G. F. & Woodward, F. I. The importance of low atmospheric CO_2 and fire in promoting the spread of grasslands and savannas. *Global Change Biol.* **9**, 973–982 (2003).
23. Whittaker, J. B. Impacts and responses at population level of herbivorous insects to elevated CO_2 . *Eur. J. Entomol.* **96**, 149–156 (1999).

24. Sala, O. E. *et al.* Biodiversity—global biodiversity scenarios for the year 2100. *Science* **287**, 1770–1774 (2000).
25. Lackner, K. S. A guide to CO_2 sequestration. *Science* **300**, 1677–1678 (2003).
26. Beerling, D. J. The impact of temperature on the northern distribution limits of the introduced species *Fallopia japonica* and *Impatiens glandulifera* in north-west Europe. *J. Biogeogr.* **20**, 45–53 (1993).
27. Baker, R. H. A. *et al.* The role of climatic mapping in predicting the potential geographical distribution of non-indigenous pests under current and future climates. *Agric. Ecosyst. Environ.* **82**, 57–71 (2000).
28. Peterson, A. T. & Vieglais, D. A. Predicting species invasions using ecological niche modeling. *BioScience* **51**, 363–371 (2001).
29. Pearson, R. G. & Dawson, T. P. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecol. Biogeogr.* **12**, 361–371 (2003).
30. Intergovernmental Panel on Climate Change. *Climate Change 2001: The Scientific Basis*. http://www.grida.no/climate/ipcc_tar/wg1/figts-22.htm (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the following for many contributions: E. Bolitho, V. Perez Canhos, D. A. L. Canhos, S. Carver, S. L. Chown, S. Fox, M. Kshatriya, D. Millar, A. G. Navarro-Sigüenza, R. S. Pereira, B. Reyers, E. Martinez-Meyer, V. Sánchez-Cordero, J. Soberón, D. R. B. Stockwell, W. Thuiller, D. A. Vieglais and K. J. Wessels, researchers involved in the Projeto de Cooperação Técnica Conservação e Manejo da Biodiversidade do Bioma Cerrado, EMBRAPA Cerrados, UnB, Ibama/DFID e RBGE/Reino Unido, and the European Bird Census Council. We thank G. Mace, J. Malcolm and C. Parmesan for valuable discussions, many funding agencies for support, and B. Orlando and others at IUCN for bringing together many of the coauthors at workshops. Comments from J. A. Pounds and S. Pimm greatly improved the manuscript.

Authors' contributions The fourth and subsequent authors are alphabetically arranged and contributed equally.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.D.T. (c.d.thomas@leeds.ac.uk).

Derivation of embryonic germ cells and male gametes from embryonic stem cells

Niels Geijsen^{1,2}, Melissa Horoschak^{1,3}, Kitai Kim^{1,3}, Joost Gribnau¹, Kevin Eggan⁴ & George Q. Daley^{1,3}

¹Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

²Center for Regenerative Medicine and Technology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA

³Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, and Division of Pediatric Hematology/Oncology, The Children's Hospital and Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA

⁴Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

Egg and sperm cells (gametes) of the mouse are derived from a founder population of primordial germ cells that are set aside early in embryogenesis. Primordial germ cells arise from the proximal epiblast, a region of the early mouse embryo that also contributes to the first blood lineages of the embryonic yolk sac¹. Embryonic stem cells differentiate *in vitro* into cystic structures called embryoid bodies consisting of tissue lineages typical of the early mouse embryo^{2,3}. Because embryoid bodies sustain blood development, we reasoned that they might also support primordial germ cell formation. Here we isolate primordial germ cells from embryoid bodies, and derive continuously growing lines of embryonic germ cells. Embryonic germ cells show erasure of the

methylation markers (imprints) of the *Igf2r* and *H19* genes, a property characteristic of the germ lineage. We show that embryoid bodies support maturation of the primordial germ cells into haploid male gametes, which when injected into oocytes restore the somatic diploid chromosome complement and develop into blastocysts. Our ability to derive germ cells from embryonic stem cells provides an accessible *in vitro* model system for studies of germline epigenetic modification and mammalian gametogenesis.

We differentiated embryonic stem (ES) cells according to our standard methods⁴, and isolated messenger RNA from whole embryoid bodies (EBs) at several time points. We used polymerase chain reaction with reverse transcription (RT-PCR) to detect expression of genes implicated in ES cell pluripotency (the POU domain transcription factor *Oct4*) and germ cell development, including *stella* and *fragilis* (*Fgls*)⁵, and a set of genes that are exclusively expressed in the germ line and are absent from somatic tissues (*Dazl*, *Piwil2*, *Rnf17*, *Rnh2*, *Tdrd1* and *Tex14*)^{6–9}. All of these genes were expressed in undifferentiated ES cells, and a subset underwent rapid extinction with EB formation (Fig. 1). *Rnh2*, *Tdrd1* and *Tex14* decreased to undetectable levels very early in EB development (day 3–4), suggesting efficient differentiation and commitment to distinct cell fates. *Stella* and *Fgls* expression declined immediately on EB formation, but low levels persisted over the course of EB differentiation.

Expression of the surface antigen SSEA1, a marker of pluripotent ES cells, also wanes on EB development, but rare SSEA1 positive

(SSEA1⁺) cells persist in differentiated EBs¹⁰. We differentiated ES cells carrying an *Oct4*-promoter-driven green fluorescent protein (GFP) reporter gene¹¹, and observed a similar overall decrease in GFP⁺ cell populations upon EB differentiation, as well as a persistent rare population of GFP⁺ cells (data not shown). Although these might represent residual undifferentiated ES cells, SSEA1 and *Oct4* expression are also features of primordial germ cells (PGCs). We thus sought to determine whether these SSEA1⁺/*Oct4*⁺ cells within EBs represented residual undifferentiated ES cells or true PGCs.

There is a lack of markers that can suitably distinguish between ES cells and PGCs; however, retinoic acid acts to rapidly differentiate ES cells while stimulating proliferation of PGCs, and can therefore be used to distinguish between these two cell populations¹². We plated ES cells or cells derived from EBs onto a mouse embryonic feeder (MEF) cell layer and cultured the cells for 7 days in the presence of 2 μ M retinoic acid (Fig. 2a). We then quantified the residual SSEA1⁺ cells. In retinoic acid-treated ES cells, greater than 99% of cells extinguished SSEA1 expression. In contrast, in retinoic acid-treated cultures of EB-derived cells, a significant percentage of SSEA1⁺ cells persisted and expanded modestly in culture, with an apparent wave of formation peaking at around day 5 (Fig. 2b). Cells that differentiated for 5 days in retinoic acid were fixed and stained for alkaline phosphatase. Very few cells positive for alkaline phosphatase persisted in the retinoic acid-treated ES cell cultures (Fig. 2c, left panel). In contrast, large colonies of cells positive for alkaline phosphatase surrounded by motile cells that resembled migratory PGCs were abundant in retinoic acid-treated cultures of EB-derived cells (Fig. 2c, right panel). This differential effect of retinoic acid strongly suggested that the SSEA1⁺ population of cells from EBs were PGCs.

To obtain conclusive evidence that the retinoic acid-resistant ES-like colonies were indeed PGCs, we analysed whether these cells manifested erasure of epigenetic imprints; this is a unique property of PGCs that is maintained in PGC-derived embryonic germ cells¹³. Imprinting of the *Igf2r* gene is determined by parental origin, with expression only from the maternal allele¹⁴. A specific region of the *Igf2r* gene has been identified, differentially methylated region 2 (DMR2), which is hypermethylated only on the maternally inherited allele¹⁵. We isolated SSEA1⁺ cells from EBs at different time points and cultured the cells for 7 days in the presence of retinoic acid. Individual retinoic acid-resistant colonies were isolated and expanded on gelatinized tissue culture plastic in the presence of leukaemia inhibitory factor (LIF), stem cell factor (SCF) and basic fibroblast growth factor; these are conditions that support the derivation of embryonic germ cells^{16,17}. We analysed the methylation status of DMR2 in independent clones of the parental ES cell line and in candidate embryonic germ clones by restriction digestion of the genomic DNA with *PvuII* and the methylation-sensitive enzyme *MluI*. Independent ES cell clones demonstrated a somatic methylation profile in which only one allele was digested (Fig. 2d, top panel, lanes 1–5). Most of the day 4 EB-derived embryonic germ cell clones displayed a similar somatic imprinting profile (Fig. 2d, top panel, lanes 6–9), with the exception of one clone in which methylation was erased (Fig. 2d, top panel, lane 10), as demonstrated by digestion of both alleles. Notably, at day 7 of EB development, 6 of 7 embryonic germ-like cells showed an unmethylated pattern (Fig. 2d, top panel, lanes 11–16), and at day 10 of EB development all embryonic germ clones had lost imprinting of the *Igf2r* gene. Similar results were obtained showing erasure of the methylation marker at the *H19/Igf2* locus (Fig. 2d, bottom panel). These data demonstrate that the EB-derived PGCs display phenotypic and biological properties of PGCs developing *in vivo*.

We then used immunomagnetic bead sorting to isolate SSEA1⁺ candidate PGCs from differentiated EBs, and used RT-PCR to analyse expression of germ-cell-specific markers. *Oct4* was

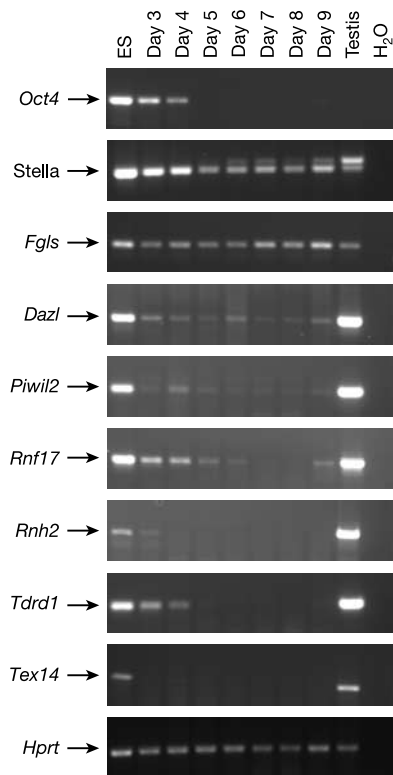


Figure 1 Expression of germ-cell-specific genes during EB development. RNA was prepared from ES cells, day 3–9 EBs and adult testis, and processed for RT-PCR. *Hprt* served as a control. *Stella* is expressed at high levels in primordial germ cells and oocytes but is almost absent from adult testis⁵. *Fgls* is expressed in the cells surrounding the germ cells⁵. *Dazl* is expressed only in germ cells in both testis and ovary^{7,29}. *Piwil2*, *Rnf17*, *Rnh2*, *Tdrd1* and *Tex14* are germ-cell-specific and expressed in the male gonad⁶.

expressed in the SSEA1⁺ population throughout EB development (Fig. 3a, left panel). *Oct4* expression was almost undetectable in the SSEA1-negative (SSEA1⁻) fraction, demonstrating the effectiveness of the SSEA1 selection (Fig. 3a, right panel). *Tex14* and *Rnh2*, which demonstrated a rapid downregulation on EB differentiation in the whole EB population, became undetectable in the purified SSEA1⁺ fraction of early EBs (days 3, 4; Fig. 3b). At day 5 expression levels of these genes rose, and by day 6 they reached a level comparable to ES cells. Although less marked than *Tex14* and *Rnh2*, expression of *Piwil2* and *Dazl* follows a similar pattern of increased expression over time in SSEA1⁺ differentiated EB-derived cell populations. This suggests that the developing EB supports a cellular environment similar to the early embryonic microenvironment in which PGCs are found.

To determine whether the PGCs arise in a defined region of the EB, we used immunohistochemistry to simultaneously visualize CD41⁺ haematopoietic cells¹⁸ and SSEA1⁺ germ cells in cryosections of 7-day-old EBs (Fig. 3c). Similar to the developing embryo, the SSEA1⁺ PGCs in the developing EB exist in close juxtaposition to the cells of the developing haematopoietic system. The colocalization of nascent blood and germ cell populations in the peripheral zones of the developing EB raises interesting questions about their clonal origins and the microenvironment that specifies their distinct cell fates.

We next investigated whether the EB-derived PGCs undergo further differentiation into functional gametes. We analysed the expression of *Sry*, a male gene that determines germ cell fate. *Sry*

expression was first detected in day 5 EBs, heralding initiation of a male germ cell developmental programme (Fig. 4a). Germ cell nuclear factor (*Gcnf*) has a role in confining *Oct4* expression to the germ line¹⁹. We observed transient expression of *Gcnf* starting at day 7 and peaking at around day 11, suggesting a temporal window during which the germ cells become fully specified. At about day 11 of EB differentiation we found a strong upregulation of acrosin and haprin, genes tightly associated with male germ cell development (Fig. 4b). Acrosin is part of the acrosomal complex transcribed as proacrosin in the diploid germ cell population²⁰, whereas haprin is a member of the RING finger-B box-coiled-coil family of transcription factors, which have a role in spermatogenesis and the formation of germ cell tumours^{21–23}. Our observation of upregulation of genes associated with male germ cell maturation prompted us to investigate the presence of stromal supporter cells. Indeed, we detected the message for the luteinizing hormone/gonadotropin receptor (LH-R) as well as müllerian inhibiting substance (MIS); these are markers of Leydig and Sertoli cells, respectively. We failed to detect expression of zona pellucida proteins Zp1 and Zp2, whereas Zp3 is expressed in ES cells and early EBs (Fig. 4b), suggesting that within the context of EB differentiation, the default programme of female gametogenesis is suppressed. Indeed when comparing the expression of two genes specific for male germ cell differentiation, *AZ1* and ret finger protein (*Rfp*), we observed exclusive expression of these genes in EBs derived from male (XY) but not female (XX) ES cells (Fig. 4c).

To investigate whether male germ cells undergo meiosis in the

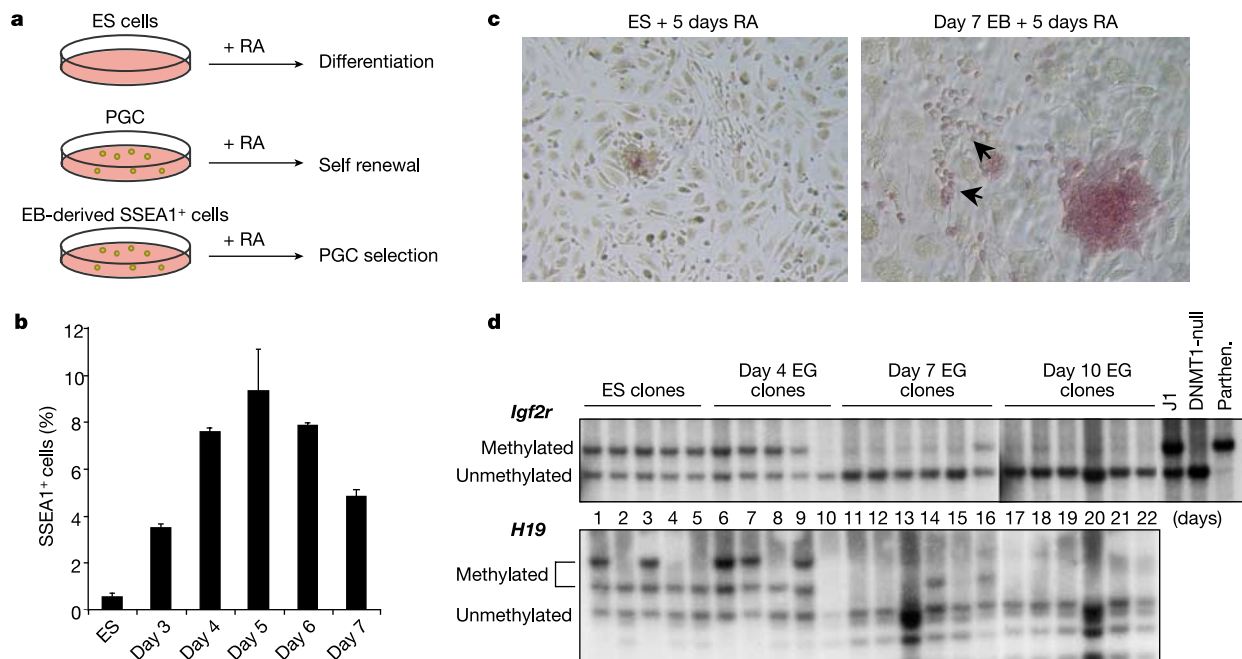


Figure 2 Development of primordial germ cells in the differentiating EB. **a**, Effects of treatment with retinoic acid (RA) on EB-derived cells. Retinoic acid differentiates ES cells (top panel) but supports PGCs¹² (middle panel). Culture of EB-derived SSEA1⁺ cells in the presence of retinoic acid selects for PGCs that retain germ cell marker expression (bottom panel). **b**, SSEA1⁺ cells were isolated from EBs of different ages by immunomagnetic beads, cultured in retinoic acid for 7 days, and SSEA1 expression was analysed by flow cytometry. The percentage of GFP⁺ ES and EB-derived cells expressing SSEA1 is plotted ($n = 3$). **c**, Alkaline phosphatase (AP) expression on ES cells (left panel) and day 7 EB-derived cells (right panel) after 5 days of culture in retinoic acid. The day 7 EB-derived cells formed large colonies of cells positive for alkaline phosphatase that resembled PGCs

on gross inspection¹⁷. Cells positive for alkaline phosphatase migrated out of the large colony (arrowheads). **d**, Imprint erasure of *Igf2r* and *H19/Igf2* loci. Top panel, DMR2 region of *Igf2r* locus. Southern analysis of genomic DNA from individual ES and EB-derived embryonic germ clones digested with *PvuII* and *MluI* is shown. Bottom panel, *H19/Igf2* locus. The same clones as in the top panel were digested with *SacI* and *HhaI*. The day of EB development at which individual embryonic germ clones were derived is indicated. Control lanes include J1 ES cells (somatic methylation profile), methylation-deficient DNMT1^{null} ES cells³⁰ and parthenogenetic (parthen.) ES cells (showing a maternal methylation pattern on both *Igf2r* alleles).

context of the EB, we immunostained cell populations with an antibody that specifically recognizes male meiotic germ cells (FE-J1²⁴), and analysed these cells for DNA content using the fluorescent DNA-binding dye Hoechst 33342. In the testes of adult mice, both FE-J1⁺ and haploid (1C) cells are readily identified (Fig. 4d, left column of top and middle panels). Analysing the FE-J1⁺ cell population for DNA content shows that this marker of male germ cells recognizes a minor diploid (2C) population of primary spermatocytes and a predominant population of cells with a haploid (1C) DNA complement, as reported²⁴ (Fig. 4d, bottom panel, left). In the EB cell suspension, haploid (1C) cells are discernable but obscured by a range of apoptotic cells with sub-2C DNA content (Fig. 4d, middle panel, right). However, when only the cells positive for the FE-J1 antibody are analysed, a distinct haploid (1C) population is observed, representing the predominant class of FE-J1-staining cells (Fig. 4d, bottom panel, right). The low proportion of antibody-positive cells (0.01%) and the relatively high ratio between diploid (2C) primary spermatocytes and haploid (1C) cells suggests that meiosis is highly inefficient in EBs. This experiment, however, demonstrates that the EB microenvironment

is permissive for male germ cell development and meiotic maturation.

We further analysed the FE-J1⁺ cells by immunofluorescence microscopy. The FE-J1 antibody recognizes the anterior acrosome on early and late pachytene spermatocytes and on round spermatids²⁴. As can be seen in Fig. 4e, FE-J1⁺ cells isolated from testes (top panel) or from day 20 EBs (bottom panel) demonstrate polarized apical perinuclear staining, indicating that the EB-derived cells have a similar morphology to testis-derived haploid cells, and possibly represent round spermatids.

Finally, we investigated the biological function of the EB-derived haploid cells and assayed their capacity to fertilize oocytes. We isolated FE-J1⁺/GFP⁺ haploid cells from day 20 EBs by flow cytometry, and performed intracytoplasmic injection into recipient oocytes. Five separate microinjection experiments performed in two independent laboratories produced comparable results. Approximately 50% of the injected oocytes supported cleavage to the 2-cell stage ($n = 125$), with 20% displaying progression to blastocysts. Figure 4f shows four representative blastocysts expressing the GFP transgene, indicating successful complementation of the oocyte genome by the EB-derived haploid cells. Furthermore, fluorescence *in situ* hybridization (FISH) using probes directed against an autosomal locus and markers on the X and Y chromosome demonstrate a normal diploid chromosome complement and expected ratios of male (XY) and female (XX) embryos (Fig. 4g). Efforts are underway to determine whether embryos arising from fertilization with EB-derived male gametes will develop normally after uterine transfer.

Germ cell development remains a largely unexplored but fascinating process of cell fate specification. Germ cells represent a privileged class of cells, given responsibility for perpetuating pluripotency and ensuring propagation of the gene pool. The genetic mechanisms that account for maintenance of pluripotency and restrict somatic differentiation are beginning to yield to molecular analyses³, but an *in vitro* model system that recapitulates germ cell specification will greatly facilitate these studies.

Recently, Schöler and colleagues reported the generation of oocytes from mouse ES cells in culture²⁵. The reported differentiation happened spontaneously over a period of nearly 50 days. In contrast to our method of ES cell differentiation into EBs, Schöler and colleagues used bulk two-dimensional differentiation on tissue culture plastic, in which both male and female lines of ES cells yielded oocytes. Given our observation of male germ cell development, we speculate that EBs may preserve more of the tissue organization reflective of the embryonic gonadal ridge, thereby enabling male germ lineage specification. In support of this notion, we detected expression of the müllerian inhibiting substance in EBs by RT-PCR. While this manuscript was under review, Noce and colleagues reported the successful derivation of male lineage germ cells from ES cells *in vitro*²⁶. Our work corroborates and extends this report by demonstrating the erasure of methylation markers at imprinted loci and the successful fertilization of oocytes by the ES-derived male gametes.

Although EBs may not reflect the precise temporal and spatial features of embryonic development, their ready derivation from ES cells *in vitro* has proven valuable for genetic studies of tissue differentiation, by linking gene deletions or ectopic transgene expression to specific cellular phenotypes. We have demonstrated the *in vitro* differentiation of ES cells into primordial germ cells, which proliferate in response to retinoic acid and give rise to embryonic germ-cell-like clones that undergo erasure of imprints. We are currently using this *in vitro* system to investigate whether EB differentiation sustains a male germ cell niche that enables the proper restoration of male imprints, thereby affording an *in vitro* system to address the genetic and biochemical mechanisms of this fundamental epigenetic modification. As demonstrated by molecu-

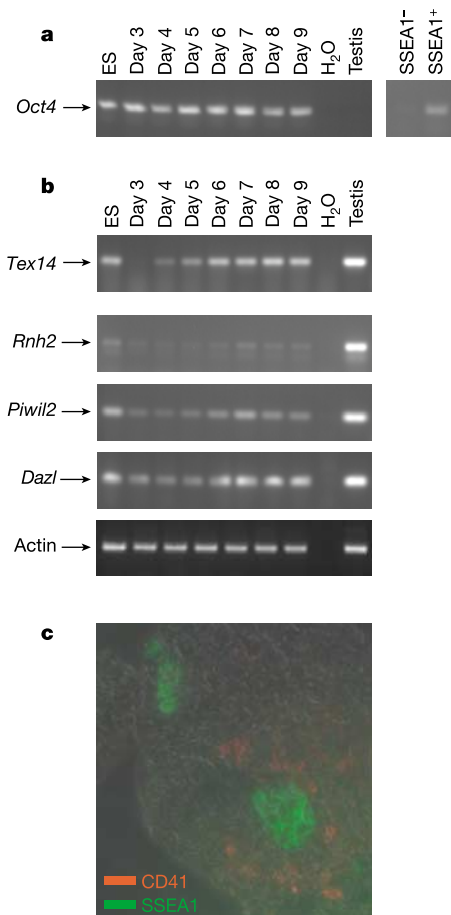


Figure 3 RT-PCR detection of germ-cell-specific genes in SSEA1⁺ cells isolated from developing EBs. **a**, Expression of *Oct4* during EB development (left panel) and in SSEA1⁺ and SSEA1⁻ fractions derived by immunomagnetic sorting. Flow cytometric analysis demonstrates >90% purity in the SSEA1⁺ fraction and >99% purity of the SSEA1⁻ fraction (data not shown). **b**, Expression of *Tex14*, *Rnh2*, *Piwil2* and *Dazl* assayed by RT-PCR. Actin served as a control. **c**, Immunohistochemistry of germ cell and haematopoietic development in day 7 EBs. EBs were cryosectioned and stained with CD41–fluorescein isothiocyanate (green) to mark early haematopoietic cells and SSEA1–PE (red) to visualize germ cell development.

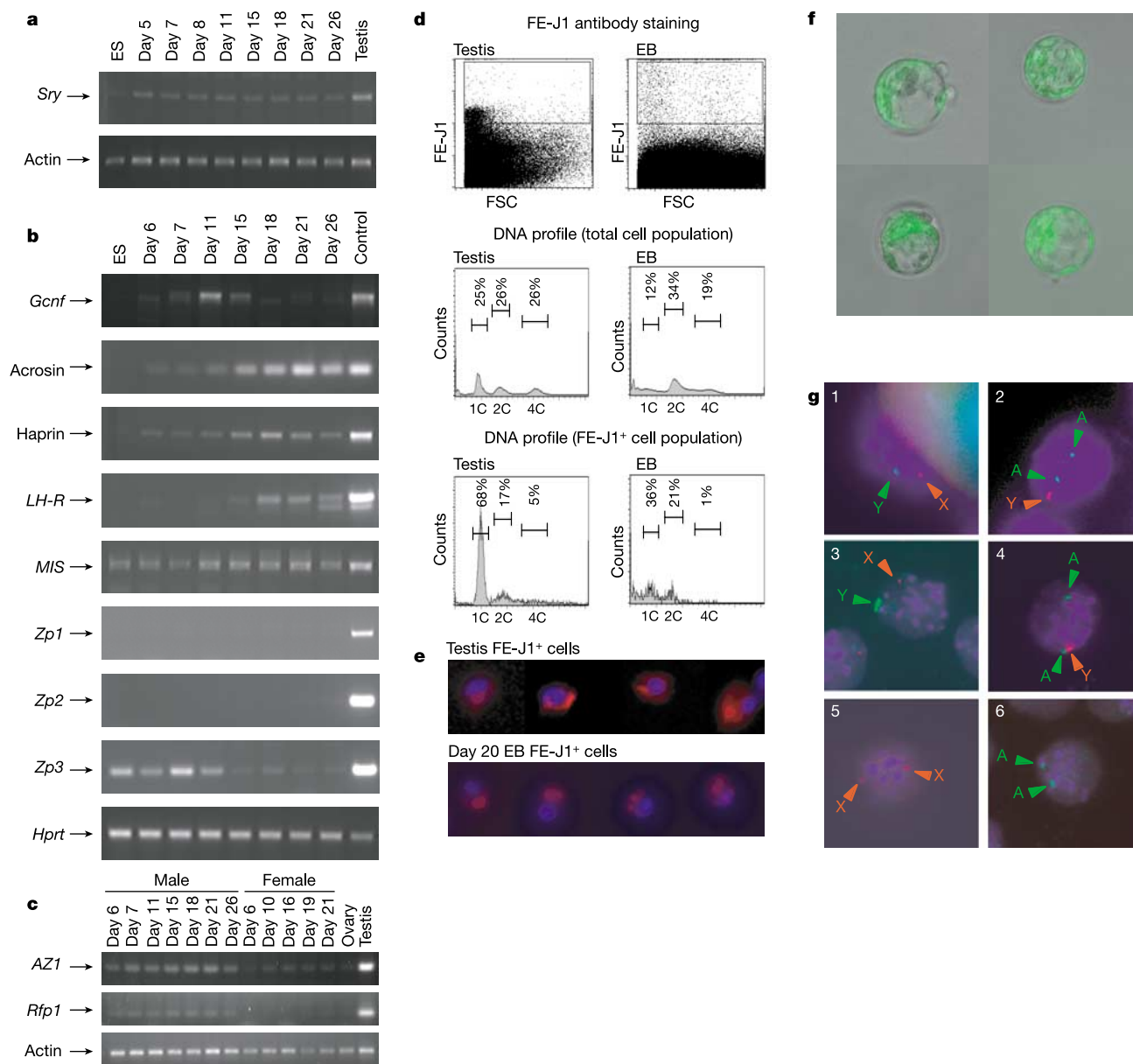


Figure 4 EBs support differentiation of haploid male germ cells that support fertilization of oocytes. **a**, Expression of *Sry* in developing EBs analysed by RT-PCR (top panel).

b, Expression of genes regulating male germ cell specification during EB development. *Gcnf* helps restrict *Oct4* expression to the germ line¹⁹. *Acrosin*, acrosomal protein with a role in oocyte interaction²⁰; *haprin*, haploid germ-cell-specific RING finger protein; *LH-R* and *MIS*, markers for Leydig and Sertoli cells, respectively; *Zp1–3*, proteins of the oocyte zona pellucida. *Hprt* served as a control. **c**, Expression of male germ-cell-specific genes in male or female EBs assayed by RT-PCR. Control RNA from ovary or testis is shown.

d, Identification of haploid male germ cells in adult testis (left) or day 13 EBs (right) by immunostaining with the FE-J1 antibody²⁴. Cells were stained with the DNA-binding dye Hoechst 33342 to reveal DNA content. 1C, haploid; 2C, G0/G1 phase; 4C, G2/M phase.

e, Immunohistochemistry of the FE-J1⁺ cells sorted in **d**. FE-J1–PE, red; nuclear Hoechst 33342, blue. **f**, Fluorescence image of blastocysts derived from intracytoplasmic injection of oocytes with EB-derived FE-J1⁺ haploid cells. **g**, FISH analysis of control ES cells and morula stage embryos, with fluorescence signals indicated by arrows. Panels 1 (X and Y chromosome markers) and 2 (autosomal (κ light chain) and Y chromosome markers) show embryos derived from intracytoplasmic oocyte injection of FE-J1⁺ haploid cells. Panels 3 (X and Y chromosome markers) and 4 (autosomal (κ light chain) and Y chromosome markers) show control male ES cells. Panels 5 (X and Y chromosome markers; note two X and absence of Y signals) and 6 (autosomal (κ light chain) and Y chromosome markers; note absence of Y signal) show control female ES cells.

lar analysis, the EBs support a programme of male germ cell differentiation, culminating in the formation of haploid cells that manifest the morphology and fertilization potential of male haploid germ cells of the round spermatid stage. Our report, together with the recent demonstration of oocyte and sperm generation from ES cells^{25,26}, signals a new realm of possibilities for investigating germ cell development, epigenetic reprogramming and germline gene modification.

Methods

Mouse strains and ES cells

C57BL/6-TGN(ACTbEGFP) mice were from Jackson Laboratories²⁷. 129SvEv mice were from Taconic. ES cells were derived from an F₁ cross between C57BL/6-TGN(ACTbEGFP)1Os and 129SvEv.

Cell culture

ES cells were maintained on irradiated MEFs in DME/15% IFS, 0.1 mM non-essential amino acids (GIBCO), 2 mM glutamine, penicillin/streptomycin (GIBCO), 0.1 mM

β -mercaptoethanol, and 1,000 U ml⁻¹ LIF (Invitrogen). For EB differentiation, ES cells were digested with trypsin, collected in EB medium (IMDM/15% IFS, 200 μ g ml⁻¹ iron-saturated transferrin (Sigma), 4.5 mM monothioglycerol (Sigma), 50 μ g ml⁻¹ ascorbic acid (Sigma) and 2 mM glutamine) and plated for 45 min to allow MEFs to adhere. Non-adherent cells were collected and plated in hanging drops at 200 cells per 30 μ l droplet in an inverted bacterial Petri dish. EBs were collected from the hanging drops at day 3 and transferred into 10 ml EB medium in slowly rotating 10 cm Petri dishes. At day 4, EBs were fed by exchanging half of their spent medium. Cells were collected by collagenase treatment and re-suspension in cell dissociation buffer (Invitrogen).

RT-PCR

RT-PCR amplifications were titrated to be within a linear range of amplification. Primers used are: *Oct4*(f) 5'-GTGATTCTCGAACCTGGCT-3', *Oct4*(r) 5'-GTCTCCAGACTCCACCTAC-3'; *stella*(f) 5'-CAGCCGTACTGTGGAGAACAGAG-3', *stella*(r) 5'-AGCCCTGGCCCTCAGACTT-3'; *Fgls*(f) 5'-TTGCTCCGCCACCATGAACCA-3', *Fgls*(r) 5'-TGAAGCACTTACAGACCGGA-3'; *Dazl*(f) 5'-GCCAGCACTCAGTCTTCATC-3', *Dazl*(r) 5'-GTTGGAGGCTGCATGTAAGT-3'; *Piwil2*(f) 5'-CCGTCATGAAGGAGAGCTCG-3', *Piwil2*(r) 5'-GGAACGACTCTGTGCTGGAT-3'; *Rnf17*(f) 5'-GACACA CAGTCTAACAGAGG-3', *Rnf17*(r) 5'-AGGACAGCAGCATCTACCTT-3'; *Rnh2*(f) 5'-CATAAGTGGCAACGAAGAG-3', *Rnh2*(r) 5'-GTTACAGGCTGTACCATCA-3'; *Tdrd1*(f) 5'-GCAGTTCTGCTCTGTCAAGG-3', *Tdrd1*(r) 5'-CAGAGCGTGAATCACA TGG-3'; *Tex14*(f) 5'-GAAGCTTGAGCAGGAGGTAG-3', *Tex14*(r) 5'-TTCAGAAGACA CAGACGCCA-3'; *Gcnf*(f) 5'-GTGGAAGACAGGACGACGA-3', *Gcnf*(r) 5'-CCTAC TGGATGATAGTGTGG-3'; *acrosin*(f) 5'-CGGAGTCTACACAGCCACCT-3', *acrosin*(r) 5'-GCATGAGTGATGAGGAGGT-3'; *haprin*(f) 5'-CCAGAACATGAGACAGAGAG-3', *haprin*(r) 5'-AGCAACTTCTGAGCATACC-3'; *Hprt*(f) 5'-GCTGGTGAAGAGGACC TCT-3', *Hprt*(r) 5'-CACAGACTAGAACACCTGC-3'; *Sry*(f) 5'-TTACAGCTGCAG TTGCTC-3', *Sry*(r) 5'-GGTCATAGAAGTCTGTGTTGC-3'; *MIS*(f) 5'-TTGGTGCTAA CCGTGGACTT-3', *MIS*(r) 5'-GCAGAGCAGCAACCAAGCGA-3'; *LH-R*(f) 5'-TGCAA CCTCTCAATCTGTC-3', *LH-R*(r) 5'-AGCGTGGCAACCAAGTGGCT-3'; *Zp1*(f) 5'-GAGTGACTGTGTGCCATAG-3', *Zp1*(r) 5'-GCCACACTGGTCTCACTACG-3'; *Zp2*(f) 5'-GCTACACATGACTCTAC-3', *Zp2*(r) 5'-GGTGACTCAGATGGCA CTC-3'; *Zp3*(f) 5'-TTGAGCAGAAGCAGTCCAGC-3', *Zp3*(r) 5'-CGGTTGCCTTGT GGATGGTC-3'; *actin*(f) 5'-ACCAACTGGGACGATATGGAGAAGA-3', *actin*(r) 5'-CTCTTTGATGTACGCACGATTC-3'.

Immunomagnetic isolation of SSEA1⁺ cells

Cells collected from EBs were incubated (30 min) with a monoclonal antibody against SSEA1 (Hybridoma bank) at 4 °C in PBS/0.5% BSA. Cells were washed twice with ice-cold PBS/0.5% BSA before addition of immunomagnetic rat anti-mouse IgM beads (Dynal), and were incubated for 1 h at 4 °C with slow rotation. Magnetic separation of SSEA1⁺ beads associated with the cells was performed according to the manufacturer's protocol.

Southern analysis

Genomic DNA was prepared from individual clones of the parental ES cell line or EB-derived embryonic germ cells. For the embryonic germ cell clones, SSEA1⁺ cells were isolated at different days of EB development and cells were grown in the presence of retinoic acid for 7 days followed by 2 days of culture without retinoic acid. Individual clones were isolated and expanded on gelatinized tissue culture plastic in the presence of LIF to remove feeder cells. Genomic DNA was isolated and digested with *PvuII* and *MluI* for the detection of *Igf2r* methylation, or with *SacI* and *HhaI* for the analysis of *H19* imprints. DNA was separated on a 0.7% agarose gel and Southern blots were generated by standard methods. Filters were hybridized with a probe covering region 2 of the *Igf2r* receptor (pPP4).

FACS analysis

Testicular and EB-derived cell suspensions were obtained by enzymatic digest of the tissue. Briefly, cells were incubated at 37 °C for 15 min with digest buffer (0.1% collagenase IV, 0.2% hyaluronidase and 50 U ml⁻¹ DNase (all Sigma)). Cells were then dissociated using cell dissociation buffer (Invitrogen), collected by centrifugation and a second digest was performed. Cell clumps were removed using a 70 μ m strainer and cells were re-suspended in ice-cold RPMI plus 0.5% FBS. Cells were incubated with FE-J1, a haploid male germ-cell-specific antibody (Hybridoma bank²⁴), for 30 min at 4 °C. The cells were washed twice with RPMI/0.5% FBS and incubated with phycoerythrin (PE)-conjugated rat anti-mouse IgM for 30 min at 4 °C. Cells were washed twice with RPMI/0.5% FBS, re-suspended in RPMI/0.5% FBS containing 2.5 μ g ml⁻¹ cytochalasin B (Sigma) and sorted on a Becton-Dickinson FACSCalibur.

Oocyte injections

Eight-ten-week-old B6D2F1/J mice (Jackson Laboratories) were used as oocyte donors. EB-derived donor cells were re-suspended in KSOM (Specialty Medium) containing 10% (w/v) polyvinyl alcohol (Sigma) and 0.01% (w/v) bovine serum albumin (Sigma). Intracytoplasmic injection into cumulus-free oocytes was carried out in H-KSOM containing 5 μ g ml⁻¹ cytochalasin B (Sigma) and 3% (w/v) sucrose at room temperature using a Nikon Eclipse TE300 microscope equipped with Narashige hydraulic micromanipulators and Hoffman modulation contrast. The injected oocytes were washed five times in KSOM to remove the cytochalasin. Reconstructed embryos were activated either in KSOM containing 10 μ M calcium ionophore A23187 (Sigma) for 5 min, followed by 2 mM 6-dimethylaminopurine (Sigma) in KSOM at 37 °C in 5% CO₂ for 3 h, or for 5 h in Ca²⁺-free medium containing 10 mM SrCl₂. Embryos were then washed five times in KSOM. Embryos were cultured in KSOM at 37 °C in 5% CO₂.

DNA-FISH

DNA-FISH was performed as described²⁸ with minor modifications. The zona was removed from embryos using acid tyrodase, and the morulae were incubated in a small drop of 0.075 M KCl on a slide for 5 min. Embryos were fixed by replacing the KCl with 3:1 methanol:acetic acid. Cells were permeabilized in 0.5% triton/PBS for 10 min, washed twice in PBS and dehydrated in 70%, 90% and 100% ethanol. Bacterial artificial chromosome (BAC) clones used for FISH analysis were from BACPAC resources and from Invitrogen. Autosomal sequences were detected with BAC RP23-20P21, which is specific for the κ light chain locus of immunoglobulin. X chromosomal sequences were detected with BAC CT7-228O4, which is specific for the *Irak1* locus. Y chromosomal sequences were detected with BAC RP24-507D23, which is specific for Y chromosomal repeats.

Received 26 June; accepted 24 November 2003; doi:10.1038/nature02247.

Published online 10 December 2003.

- Lawson, K. A. & Hage, W. J. Clonal analysis of the origin of primordial germ cells in the mouse. *Ciba Found. Symp.* **182**, 68–84, 84–91 (1994).
- Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. The *in vitro* development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J. Embryol. Exp. Morphol.* **87**, 27–45 (1985).
- Leahy, A., Xiong, J. W., Kuhnert, F. & Stuhlmann, H. Use of developmental marker genes to define temporal and spatial patterns of differentiation during embryoid body formation. *J. Exp. Zool.* **284**, 67–81 (1999).
- Kyba, M., Perlingeiro, R. C. & Daley, G. Q. HoxB4 confers definitive lymphoid-myeloid engraftment potential on embryonic stem cell and yolk sac hematopoietic progenitors. *Cell* **109**, 29–37 (2002).
- Saitou, M., Barton, S. C. & Surani, M. A. A molecular programme for the specification of germ cell fate in mice. *Nature* **418**, 293–300 (2002).
- Wang, P. J., McCarrey, J. R., Yang, F. & Page, D. C. An abundance of X-linked genes expressed in spermatogonia. *Nature Genet.* **27**, 422–426 (2001).
- Cooke, H. J., Lee, M., Kerr, S. & Ruggiu, M. A murine homologue of the human DAZ gene is autosomal and expressed only in male and female gonads. *Hum. Mol. Genet.* **5**, 513–516 (1996).
- Seboun, E. *et al.* Gene sequence, localization, and evolutionary conservation of DAZLA, a candidate male sterility gene. *Genomics* **41**, 227–235 (1997).
- Moore, F. L. *et al.* Human Pumilio-2 is expressed in embryonic stem cells and germ cells and interacts with DAZ (Deleted in AZoospermia) and DAZ-like proteins. *Proc. Natl Acad. Sci. USA* **100**, 538–543 (2003).
- Zandstra, P. W., Le, H. V., Daley, G. Q., Griffith, L. G. & Lauffenburger, D. A. Leukemia inhibitory factor (LIF) concentration modulates embryonic stem cell self-renewal and differentiation independently of proliferation. *Biotechnol. Bioeng.* **69**, 607–617 (2000).
- Bortvin, A. *et al.* Incomplete reactivation of Oct4-related genes in mouse embryos cloned from somatic nuclei. *Development* **130**, 1673–1680 (2003).
- Koshimizu, U., Watanabe, M. & Nakatsuji, N. Retinoic acid is a potent growth activator of mouse primordial germ cells *in vitro*. *Dev. Biol.* **168**, 683–685 (1995).
- Labosky, P. A., Barlow, D. P. & Hogan, B. L. Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (*Igf2r*) gene compared with embryonic stem (ES) cell lines. *Development* **120**, 3197–3204 (1994).
- Barlow, D. P., Stoger, R., Herrmann, B. G., Saito, K. & Schweifer, N. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the *Time* locus. *Nature* **349**, 84–87 (1991).
- Stoger, R. *et al.* Maternal-specific methylation of the imprinted mouse *Igf2r* locus identifies the expressed locus as carrying the imprinting signal. *Cell* **73**, 61–71 (1993).
- Matsui, Y., Zsebo, K. & Hogan, B. L. Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841–847 (1992).
- Resnick, J. L., Bixler, L. S., Cheng, L. & Donovan, P. J. Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359**, 550–551 (1992).
- Mikkola, H. K., Fujiwara, Y., Schlaeger, T. M., Traver, D. & Orkin, S. H. Expression of CD41 marks the initiation of definitive hematopoiesis in the mouse embryo. *Blood* **101**, 508–516 (2003).
- Fuhrmann, G. *et al.* Mouse germline restriction of Oct4 expression by germ cell nuclear factor. *Dev. Cell* **1**, 377–387 (2001).
- Wassarman, P. M. Mouse gamete adhesion molecules. *Biol. Reprod.* **46**, 186–191 (1992).
- Cao, T., Shannon, M., Handel, M. A. & Etkin, L. D. Mouse ret finger protein (rfp) proto-oncogene is expressed at specific stages of mouse spermatogenesis. *Dev. Genet.* **19**, 309–320 (1996).
- Ogawa, S. *et al.* Molecular cloning of a novel RING finger-B box-coiled coil (RBCC) protein, *terf*, expressed in the testis. *Biochem. Biophys. Res. Commun.* **251**, 515–519 (1998).
- Tezel, G., Nagasaka, T., Shimono, Y. & Takahashi, M. Differential expression of RET finger protein in testicular germ cell tumors. *Pathol. Int.* **52**, 623–627 (2002).
- Fenderson, B. A., O'Brien, D. A., Millette, C. F. & Eddy, E. M. Stage-specific expression of three cell surface carbohydrate antigens during murine spermatogenesis detected with monoclonal antibodies. *Dev. Biol.* **103**, 117–128 (1984).
- Hubner, K. *et al.* Derivation of oocytes from mouse embryonic stem cells. *Science* **300**, 1251–1256 (2003).
- Toyooka, Y., Tsunekawa, N., Akasu, R. & Noce, T. Embryonic stem cells can form germ cells *in vitro*. *Proc. Natl Acad. Sci. USA* **100**, 11457–11462 (2003).
- Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. 'Green mice' as a source of ubiquitous green cells. *FEBS Lett.* **407**, 313–319 (1997).
- Selig, S., Okumura, K., Ward, D. C. & Cedar, H. Delineation of DNA replication time zones by fluorescence *in situ* hybridization. *EMBO J.* **11**, 1217–1225 (1992).
- Reijo, R. *et al.* Mouse autosomal homolog of DAZ, a candidate male sterility gene in humans, is expressed in male germ cells before and after puberty. *Genomics* **35**, 346–352 (1996).
- Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).

Acknowledgements The authors thank S. L. Opitz, N. Watson and E. Dikovskaia for technical support; K. Hochedlinger and T. Holm for assistance with intracytoplasmic oocyte injection; and W. Lensch for critical revisions of the manuscript. This work was supported by grants from the National Institutes of Health, the National Science Foundation Biotechnology Process Engineering Center, and the Dutch Cancer Society 'Koningin Wilhelmina fonds'. G.Q.D. is a Birnbaum Scholar of the Leukemia and Lymphoma Society of America. J.G. was sponsored by the Human Frontiers Science Foundation. K.E. is a Junior Fellow in the Harvard Society of Fellows.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.Q.D. (george.daley@childrens.harvard.edu).

T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases

Thorsten R. Mempel, Sarah E. Henrickson & Ulrich H. von Andrian

The CBR Institute for Biomedical Research, Department of Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, Massachusetts 02115, USA

Primary T-cell responses in lymph nodes (LNs) require contact-dependent information exchange between T cells and dendritic cells (DCs). Because lymphocytes continually enter and leave normal LNs, the resident lymphocyte pool is composed of non-synchronized cells with different dwell times that display heterogeneous behaviour in mouse LNs *in vitro*^{1–3}. Here we employ two-photon microscopy *in vivo* to study antigen-presenting DCs and naive T cells whose dwell time in LNs was synchronized. During the first 8 h after entering from the blood, T cells underwent multiple short encounters with DCs, progressively decreased their motility, and upregulated activation markers. During the subsequent 12 h T cells formed long-lasting stable conjugates with DCs and began to secrete interleukin-2 and interferon- γ . On the second day, coinciding with the onset of proliferation, T cells resumed their rapid migration and short DC contacts. Thus, T-cell priming by DCs occurs in three successive stages: transient serial encounters during the first activation phase are followed by a second phase of stable contacts culminating in cytokine production, which makes a transition into a third phase of high motility and rapid proliferation.

Naive T cells recirculate continually between the blood and LNs to search for antigen⁴. The intranodal encounter of a peptide-major histocompatibility complex (MHC) complex that is recognized by a T-cell antigen receptor (TCR) will only result in full-fledged T cell activation upon co-stimulation provided by mature DCs. These professional antigen-presenting cells collect antigens in peripheral tissues and migrate to LNs through lymph vessels. T-cell priming by DCs induces activation markers, cytokine secretion, and proliferation. Several reports have analysed the dynamics of T-cell–DC interactions in excised LNs, but the methods and results have been variable and it is unknown how the absence of lymph and blood flow or innervation influences T-cell–DC interactions^{1–3,5}. There is therefore still no comprehensive description of what happens in a truly physiological setting when naive T cells enter LNs that contain antigen-presenting DCs.

We have used two-photon microscopy⁶ to study lymphocyte migration and interactions with DCs within popliteal LNs of anaesthetized mice. Our preparation preserved physiological blood and lymph flow⁷, whereas, in our hands, lymph flow was

compromised when we attempted to adapt the well-established inguinal lymph-node preparation to two-photon imaging^{8,9} (not shown). Recipient mice received footpad injections of fluorescent DCs, which entered lymph vessels and accumulated in the popliteal LN during the following day. Differentially tagged TCR transgenic CD8⁺T cells were injected intravenously 18 h later (Fig. 1a). T cells homed rapidly through high endothelial venules (HEVs) into popliteal LNs⁷, where they constituted 1–2% of all CD8⁺ cells 2 h after injection. At this point, further lymphocyte homing was blocked by the injection of anti-L-selectin¹⁰. This ensured that all imaged T cells had entered popliteal LNs during the initial 2-h window and enabled us to study synchronized resident cells during the subsequent 2 days.

Initially, we examined the phenotype of injected DCs that migrated to draining LNs (Fig. 1b). Immature splenic CD11c⁺ DCs from CD45.1⁺ donors were injected into CD45.2⁺ congenic recipients. Because immature DCs express little or no CCR7—a chemokine receptor required for DC migration into lymphatics and within LNs¹¹—we co-injected lipopolysaccharide (LPS; 10 ng), which induces DC maturation and CCR7 expression¹². Consequently, DCs recovered from popliteal LNs were mature and either CD11b⁺CD8 α [–] or CD11b[–]CD8 α ^{low/–}. CD8 α ^{high} input DCs were rarely detected in LNs. This might reflect an inability of CD8 α ^{high} DCs to migrate to LNs, or the downregulation of CD8 α during transit.

To examine DCs *in situ*, we administered footpad injections of red fluorescent DCs and analysed their behaviour within popliteal LNs. After 20 h a fraction of migrated DCs (about 20–50%) resided in the subcapsular sinus and the superficial cortex. The remainder congregated in distinct regions within the deep cortex, where they localized with homed T cells (Supplementary Fig. S1a). A reason for the non-uniform distribution of DCs in the T cell area became apparent after fluorescein isothiocyanate–dextran injection to delineate blood vessels. In line with recent observations¹³, many DCs lined up around HEVs in strategic positions to interact with newly homed T cells (Supplementary Fig. S1b; Supplementary Information 2, 3). This distribution pattern became less apparent at day 2 or 3 after injection (Supplementary Information 4).

One day after injection, DCs in T-cell areas were remarkably motile, whether they presented antigen to T cells or not (Fig. 1c–g; Supplementary Information 5). The median three-dimensional (3D) instantaneous velocity was 6.6 $\mu\text{m min}^{-1}$ (Fig. 1e), which is in good agreement with measurements of two-dimensional velocities in explanted LNs³. Individual DCs followed random paths without apparent directional bias. Even sessile DCs constantly extended and retracted dendrites and pseudopods. The motility of both antigen-pulsed and control DCs was highest between 2 and 8 h after T-cell transfer (20–26 h after DC injection) and decreased over time (Fig. 1f, g; Supplementary Information 5). This progressive decrease in motility might reflect DC 'exhaustion', which has been proposed to occur after LPS activation¹⁴.

Migratory dynamics in LNs were closely dependent on physiological conditions: T cells and DCs stopped migrating and assumed a round shape within minutes after cardiovascular arrest (Supplementary Information 6), even with a tissue temperature maintained at 36 °C, a requirement for interstitial T-cell migration². In living animals without DC injection, intranodal T cells moved rapidly, reaching peak 3D velocities of 40 $\mu\text{m min}^{-1}$ (Fig. 2a–c; Supplementary Information 7). In LNs draining DC-injected footpads, 3D velocities were equivalent (Fig. 2c, d) but T cells turned at steeper angles and covered a smaller volume of the paracortex, resulting in decreased motility coefficients, even when DCs carried no antigen (Fig. 2e, f). Mean T-cell displacement plots in the antigen-containing LNs revealed a plateau at about 15 μm displacement (Fig. 2e), indicating confined motility¹⁵. Thus, DC plus LPS injections induced LN paracortex partitioning into smaller compartments in which incoming T cells were retained, perhaps owing

to motility-attenuating mediators released by DCs and/or other cells, or adhesive contacts between T cells and intranodal structures or confined clusters of DCs¹⁶.

To address the latter possibility, we examined T-cell migratory behaviour and the kinetics of T-cell–DC interactions in different LN areas. Superficial T cells (less than 100 μm below the capsule) were less motile than those in the deep T-cell area, whether DCs were injected or not. There was additional heterogeneity in the deep T-cell zone of LNs that drained DC injection sites; T cells in areas with high DC density were less motile than in DC-poor areas. We therefore analysed only regions that were more than 150 μm below the LN capsule and contained comparably high DC densities. Here, T-cell motility depended on the presence of antigen and T-cell

dwell time (Fig. 2g). Although instantaneous 3D velocities remained constant in LNs that contained control DCs, T cells migrated more slowly after encountering antigen-bearing DCs for 8–20 h. In contrast, T-cell migration after antigen exposure for 2 or 44 h was similar to T-cell migration in the presence of control DCs.

When T cells encountered DCs shortly after homing (2–4 h after transfer), they formed predominantly brief contacts before detaching to interact with other DCs nearby (Fig. 3a–d; Supplementary Information 8). Contacts lasted slightly longer with antigen-loaded DCs (mean $\pm$ s.e.m. 5.9 ± 0.8 min; median 3.8 min) than with control DCs (4.1 ± 0.4 min; median 2.5 min; $P < 0.05$). At 5–8 h, contacts lasted noticeably longer (median 3.8 min without antigen; 7.3 min with antigen) but rarely exceeded 30 min (Fig. 3e). A

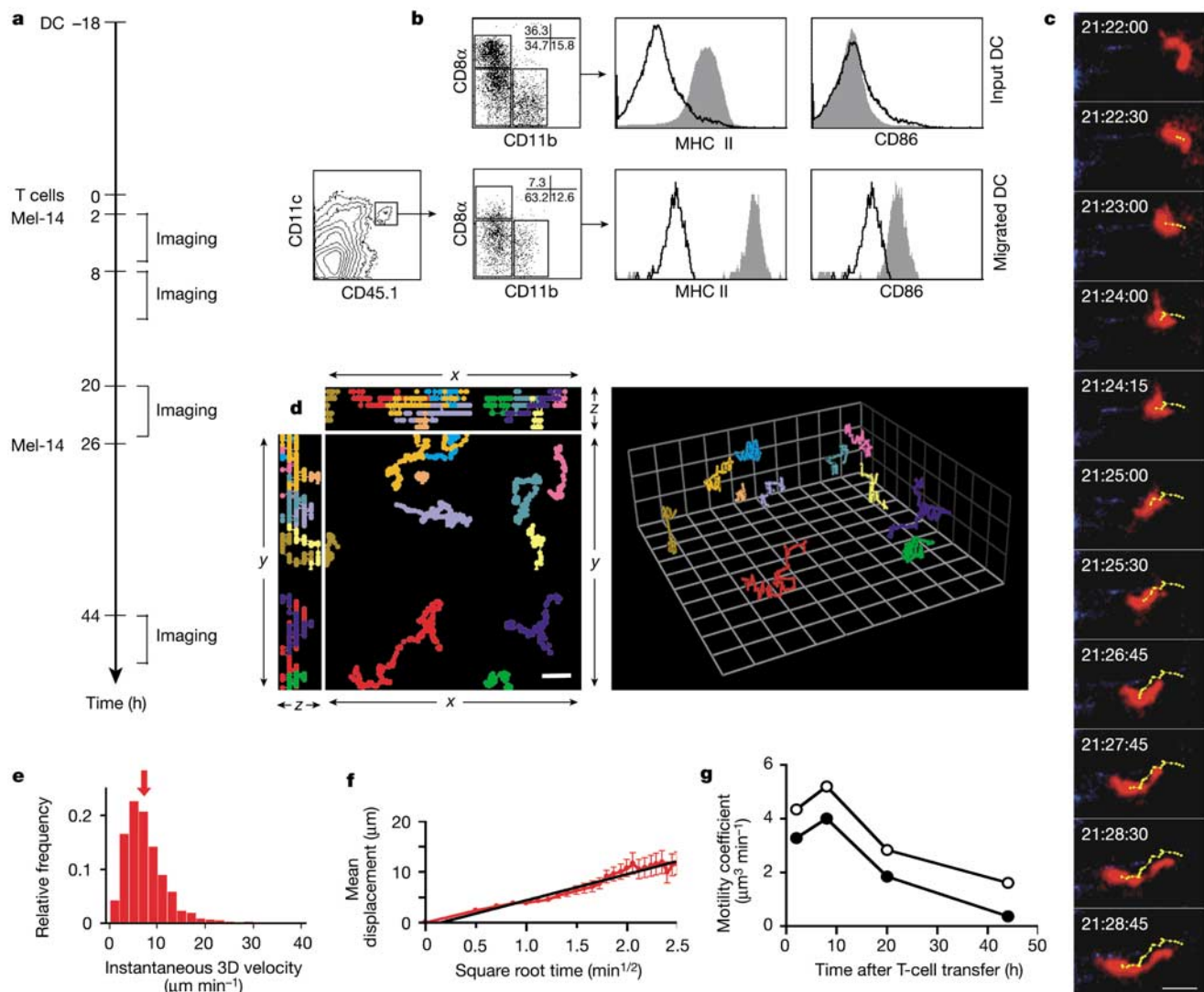


Figure 1 Surface phenotype and migratory properties of DCs in popliteal LNs.

a, Experimental protocol. Antigen-pulsed or unpulsed DCs were injected into recipient footpads 18 h before adoptive transfer of T cells. Animals received anti-L-selectin monoclonal antibody Mel-14 2 h and 26 h later, and popliteal LNs were imaged at different times thereafter. **b–f**, Data from unpulsed DCs at 2–3 h after T-cell injection. **b**, CD11c⁺CD45.1⁺ donor DCs were examined by flow cytometry before injection into the footpad of CD45.2⁺ recipients (input DC) and after their migration into LNs (migrated DC) for subset composition (CD8 α , CD11b) and maturation state (MHC class II, CD86). **c**, Intravital two-photon micrographs of a representative intranodal CMTMR-labelled DC (see also Supplementary Information 5). The migratory path of the DC was illustrated by

tracking the 3D centroid (yellow dots) in successive image stacks. Time after DC injection is shown. The mean 3D velocity of this cell was $8.4 \mu\text{m min}^{-1}$. Scale bar, 20 μm . **d**, Representative tracks of migrating DCs depicted in x–y, x–z and y–z views (left; scale bar, 20 μm) and displayed in 3D (right; grid squares are $20.6 \mu\text{m} \times 20.6 \mu\text{m}$). **e**, Frequency histograms of 3D instantaneous velocities ($n = 840$). Arrow indicates median ($6.6 \mu\text{m min}^{-1}$). **f**, Mean displacement plot for DCs shown in **e**. The straight line denotes the regression function of the initial, linear segment of curve ($r^2 = 0.95$). The motility coefficient M was calculated from the slope of the regression line as $M = x^2/6t$, where x is the displacement at time t (ref. 2), and was $4.35 \mu\text{m}^3 \text{min}^{-1}$. **g**, Time course of motility coefficients for unpulsed (open circles) and antigen-pulsed (filled circles) DCs.

marked change in T-cell–DC conjugation was apparent when LNs were imaged at 8–12 h (Supplementary Information 9). Most conjugates with antigen-presenting DCs lasted longer than 30 min (86%), and the majority (64%) remained stable throughout the 60 min observation period. We followed single conjugates over up to 3 h, indicating that these interactions can be sustained for a substantial period. Prolonged interactions were still apparent at 20–26 h, but about 70% of the contacts had returned to short encounter mode. One day later (44–48 h), T cells engaged in brief contacts only (Supplementary Information 4).

These observations indicate that naive CD8 T cells interacted with antigen-presenting DCs in three sequential stages. Phase one constituted the first approximately 8 h after T-cell entry into LNs and was characterized by short encounters of rapidly migrating cells

with numerous DCs, preferentially in the vicinity of HEVs. By contrast, *in vitro* stimulation of TCRs on T cells migrating on two-dimensional surfaces provides an immediate stop signal followed by assembly of the so-called supramolecular activation cluster or immunological synapse at the contact site^{17–20}. The formation of mature synapses between naive T cells and antigen-presenting cells requires about 30 min (ref. 20). It is controversial whether and when immunological synapses form *in vivo* and whether they are required for T-cell activation. For example, T-cell–DC interactions in collagen matrices last only minutes but induce T-cell activation²¹. Given the short duration of contacts during phase one, it is unlikely that mature immunological synapses could have formed. Nevertheless, in contrast with LNs that did not receive DCs, T cells in DC-containing LNs upregulated CD44 and CD69 during the first 8 h

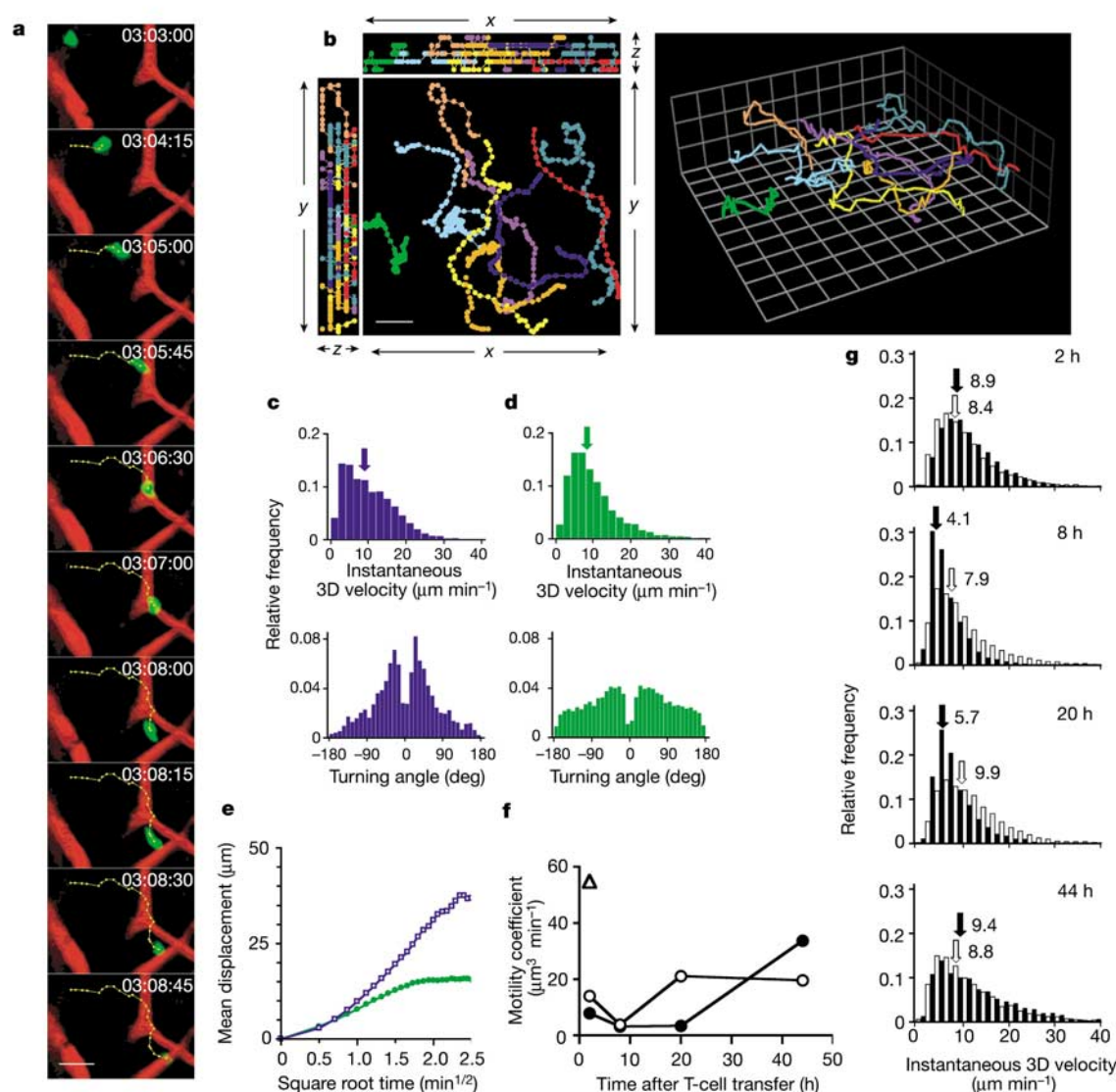


Figure 2 Effect of antigen on T-cell motility in LNs. **a**, Intranasal T cell migrating between tetramethylrhodamine β -isothiocyanate–dextran-filled capillaries in the absence of transferred DCs (see also Supplementary Information 7). Scale bar, 20 μm . **b**, Two-dimensional (left; scale bar, 20 μm) and 3D (right; grid squares are 15.4 $\mu\text{m} \times 15.4 \mu\text{m}$) depiction of representative T-cell tracks. **c**, **d**, Representative frequency histograms of 3D instantaneous velocities (upper panels) and turning angles (lower panels) of T cells migrating in the absence (**c**) or presence (**d**) of LPS-matured DCs. Arrows in the upper panels indicate medians (8.9 $\mu\text{m min}^{-1}$ in **c**; 8.3 $\mu\text{m min}^{-1}$ in **d**). **e**, Mean displacement

plots of T cells in the absence (blue) or presence (green) of LPS-matured DCs in LNs. **f**, Time course of T-cell motility coefficients in the presence of control DCs (open circles) and antigen-bearing DCs (filled circles). Triangle denotes T-cell motility coefficient without DC injection. **g**, 3D instantaneous velocity histograms of T cells in representative LNs at various times after adoptive transfer in the presence of antigen-pulsed (filled bars) or control (open bars) DCs. Medians are indicated by arrows; values shown are in units of $\mu\text{m min}^{-1}$.

(Fig. 3f, g). Thus, the brief interactions with DCs were sufficient for T-cell activation, at least at the level of these early activation markers. However, the interleukin (IL)-2 receptor, CD25, was only marginally increased, and isolated LN T cells did not spontaneously produce IL-2 or interferon (IFN)- γ after 8 h (not shown). Even when T cells were restimulated for 3 h, few cells secreted cytokines, and only at low levels (Fig. 3h).

Phase two lasted from about 8 h until about 24 h after T-cell transfer. This phase was dominated by T-cell–DC conjugates that remained stable for more than 1 h. The low T-cell velocities during this stage (Fig. 2g) reflect, in part, the slow migration of DCs that dragged interacting T cells along with them (Supplementary Information 10). Thus, active T-cell migration was even more depressed than was indicated by velocity measurements. Many T

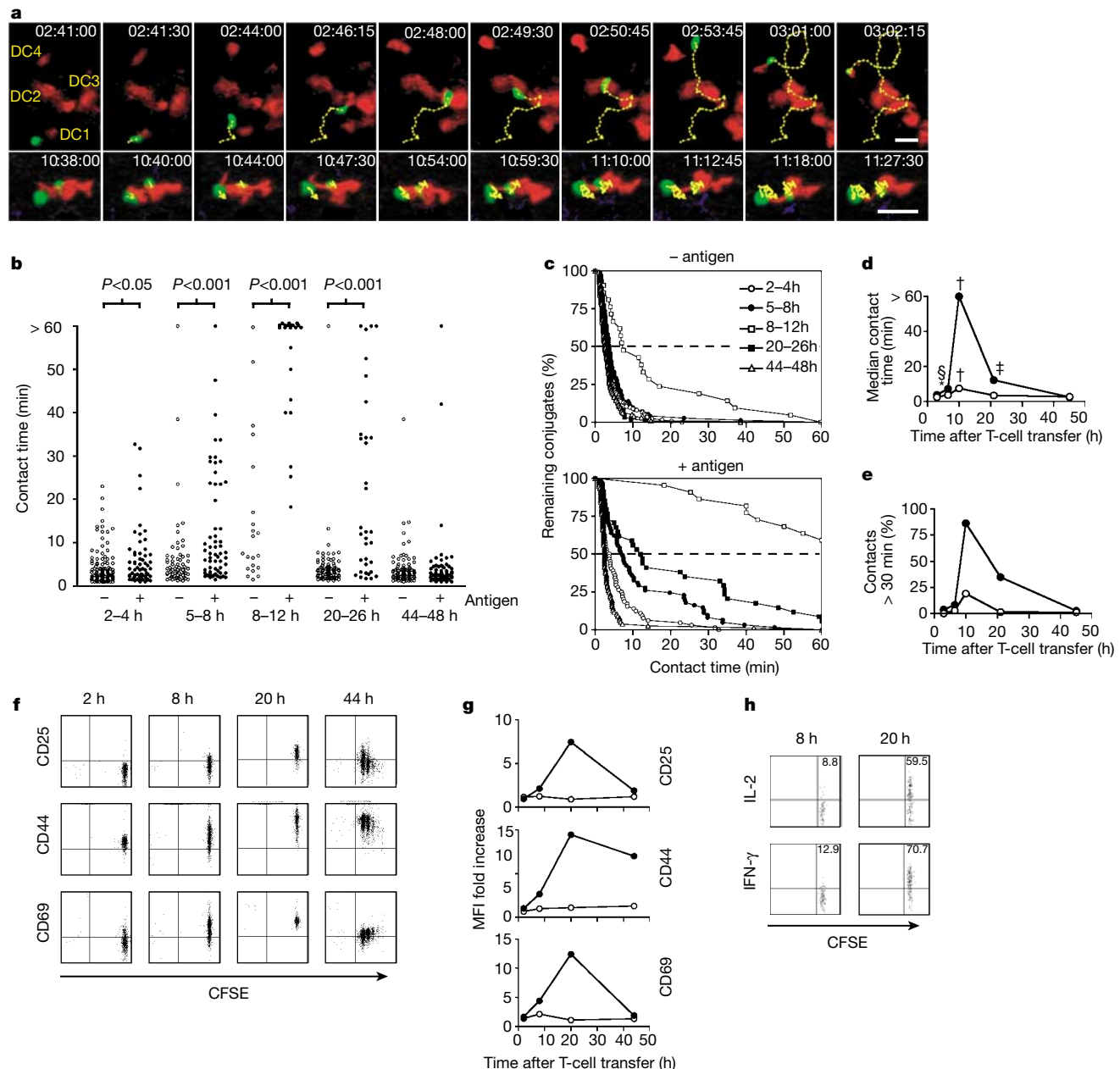


Figure 3 Progressive changes in T-cell–DC interaction dynamics indicate three phases of antigen recognition. **a**, Upper row: a representative T cell (green) early after LN entry interacts sequentially with four different antigen-pulsed DCs (red). Yellow dots indicate tracks of the 3D centroid. Lower row: at more than 10 h after homing into the LN, two T cells are stably bound to an antigen-bearing DC. The time after T-cell transfer is shown in each panel. Scale bars, 20 μ m. **b**, **c**, Absolute (**b**) and cumulative (**c**) contact duration between T cells and DCs in the presence and absence of antigen. **d**, Median contact time between T cells and DCs in the presence (filled circles) and absence (open circles) of antigen. Asterisk, $P < 0.05$ compared with 2 h; dagger, $P < 0.05$ compared with 2, 5, 20 and 44 h; double dagger, $P < 0.05$ compared with 44 h; section sign, $P < 0.05$

compared with 2 and 44 h. **e**, Frequency of T-cell–DC contacts lasting more than 30 min in the presence (filled circles) and absence (open circles) of antigen. **f**, Representative two-colour flow cytometry plots of CD25, CD44 and CD69 expression on CFSE-labelled T cells (gated on CD8⁺ MHC tetramer-positive) after different exposure times to antigen-presenting DCs. CFSE dilution indicates cell division. **g**, Time course of T-cell activation marker expression ratio in LNs containing antigen-pulsed (filled circles) or control (open circles) DCs versus contralateral LNs containing no DCs (set to 1). **h**, Stimulated secretion of IL-2 (top) and IFN- γ (bottom) by gated P14 TCR-specific MHC-tetramer-positive T cells harvested from LNs containing antigen-presenting DCs at 8 h (left) or 20 h (right) after T-cell transfer.

cells spontaneously produced IL-2 (18.6%) and IFN- γ (18.8%) at 20 h, and most secreted cytokines after brief restimulation (Fig. 3h). Almost all T cells expressed abundant activation markers, including CD25, but no cells had divided. The longevity of T-cell–DC interactions in phase two is consistent with the formation of a mature synapse-like contact zone²⁰. Recent studies on CD4⁺ T-cell interactions with DCs in excised LNs found that at 23 h after T-cell transfer the sialomucin CD43 was excluded from the contact zone of interacting cells, implying the formation of a synapse-like interface¹. However, this study imaged T cells in the superficial cortex where they formed only one-to-one associations with DCs. In the present analysis of the deep paracortex we frequently observed several T cells nestling against the same DC (Supplementary Information 9).

The formation of CD8⁺ T-cell clusters around single DCs was also detected in explanted LNs that were imaged 20 h after T-cell transfer³. Consistent with T-cell–DC behaviour during phase two *in vivo* was the observation that conjugates in excised LNs were stable, whereas interactions without antigen were short-lived³. Clustering of antigen-specific CD4⁺ T cells around DCs in LNs was also observed histologically^{22,23}. Because these observations indicated that CD4⁺ and CD8⁺ T cells might behave similarly *in vivo*, we performed experiments with DO11.10 T cells expressing a MHC class-II restricted TCR²⁴. T cells formed stable clusters with peptide-pulsed DCs when observed at 10 h, but engaged preferentially in short contacts at 3 and 47 h after injection (Supplementary Information 11), indicating that the three-phase model applies also to CD4⁺ T cells.

Phase three began 1 day after homing. T cells dissociated from DCs, migrated rapidly and proliferated vigorously. Dividing cells expressed less CD25 and CD69, but remained CD44^{high} (Fig. 3g). Although it is likely that some T cells left the LN via efferent lymph vessels after day 2, the actual number of intranodal TCR transgenic cells expanded because T cells had divided up to five times during the preceding day and kept proliferating thereafter, resulting in the dilution of carboxyfluorescein succinimidyl ester (CFSE) to undetectable concentrations (not shown). T cells still engaged in brief contacts with DCs at 44–48 h, but the relevance of these interactions is uncertain, because it is possible that DCs retained little or no peptide antigen at this time point (more than 60 h after DC injection).

Even when DCs did not present antigen, T cells underwent prolonged interactions during the 8–12-h interval, whereas earlier and later contacts were short-lived (Fig. 3b, c). This indicates that recirculating T cells follow a three-phase itinerary by default when they survey lymphoid tissues. Without antigen, phase two is abbreviated, and although T-cell–DC contacts are shorter than in the presence of antigen, they are still longer than in phase one. T-cell contacts with control DCs did not result in detectable changes in activation markers (Fig. 3g) or cytokine production (not shown). Nevertheless, TCR signalling probably occurred, because naive T cells in LNs contain partly phosphorylated CD3 ζ , even in the absence of antigen, whereas blood-borne T cells do not carry this indicator of early TCR signalling²⁵. Antigen-independent TCR stimulation in LNs depends on self-MHC molecules and facilitates T-cell responses to foreign antigens²⁵.

The intermittent period of antigen-independent, prolonged contacts might be indicative of physiological self-MHC recognition. If T cells fail to perceive an activating signal, even during this phase when they scrutinize DCs most intensely, they might be prompted to seek another lymphoid tissue. The frantic migration of T cells during phase three might signify their search for an exit into lymph sinuses. Indeed, most lymphocytes appear in thoracic duct lymph at 24–48 h after intravenous bolus injection⁴. Accordingly, we found that after anti-L-selectin treatment the number of homed T cells in LNs that did not contain antigen-pulsed DCs remained initially constant but there was a 60% decrease between 20 and 44 h (not shown). Recirculating lymphocytes therefore spend just enough

time in LNs to go through all three phases before leaving through efferent lymphatics.

It will be important to determine the mechanisms that cause the transition from one phase to the next. Although this could involve changes in both T cells and DCs, several observations indicate that adaptive T-cell responses might be chiefly responsible. T-cell velocity and motility were dependent on antigens and paralleled the differential dynamics of conjugate formation, whereas DC motility gradually declined by default and was minimally affected by antigen presentation. During the transition period before and after phase two, numerous DCs interacted stably with some T cells and simultaneously engaged in brief contacts with others (Supplementary Information 10). If contact duration was determined by DCs, then individual DCs should support either brief encounters or stable conjugation, but not both. Moreover, when T-cell dwell time was not synchronized by anti-L-selectin, stable contacts were prominent at 10 h after T-cell injection, but short-lasting contacts were more frequent than after treatment with antibody, presumably owing to recent immigrants displaying phase-one-like behaviour (Supplementary Information 11, 12).

How do intranodal T cells keep track of time and what are the immunological consequences of each interactive phase? *In vitro*, T cells can integrate multiple brief signals to reach a threshold for full-fledged activation^{21,26}. Conceivably, this is how T cells gather information from DCs during phase one. However, the ability of *in vitro* primed CD8⁺ T cells to secrete IL-2 and to exert effector activities after adoptive transfer depends on the formation of stable T-cell–DC conjugates and the duration of T-cell stimulation, respectively^{27,28}. Thus, phase two interactions might also be important. The three-stage time course of T-cell priming by DCs documented here provides a roadmap for testing these concepts and the underlying mechanisms *in situ*. □

Methods

Mice

Male Balb/c, C57BL/6 and congenic C57BL/6 (CD45.1) mice were purchased from Jackson Laboratories and used at the age of 6–12 weeks. DO11.10 and P14 mice, which carry a transgenic TCR specific for OVA_{323–339} in I-A^d and LCMV gp33–41 in H-2D^b, respectively, were obtained from Jackson Laboratories^{24,29}. All experiments were performed in accordance with National Institute of Health guidelines and approved by the Committees on Animal Care and Use of both Harvard Medical School and the CBR Institute for Biomedical Research.

Reagents

OVA_{323–339} peptide (ISQAVHAHAHAEINEAGR) was a gift from H. Ploegh. LCMV gp33–41 peptide (KAVYNEATC) was purchased from Biosource. Anti-L-selectin monoclonal antibody Mel-14 was grown and purified according to standard procedures. All other monoclonal antibodies were from BD Pharmingen.

Cells

DCs were immunomagnetically purified (97% CD11c⁺) from spleens of donor mice that had been implanted with a Flt-3L secreting mouse melanoma cell line as described³⁰. CD4⁺ and CD8⁺ T cells from LNs and spleens of DO11.10 and P14 mice were purified by negative immunomagnetic cell sorting (Miltenyi Biotec). Purity was typically more than 90%.

Flow cytometry

Phenotypic characterization of DCs and T cells was performed on a FACSCalibur (Becton Dickinson). The secretion of IL-2 and IFN- γ was measured by surface capture (Miltenyi Biotec). For restimulation, T cells were incubated for 3 h with 10 μ g ml^{−1} gp33–41 peptide at 10⁷ cells ml^{−1} in medium at 37 °C containing 10% murine serum. For the measurement of spontaneous cytokine release, antigenic peptide was omitted. Phycoerythrin-labelled gp33–41-MHC class I tetramer (H-2D^b-PE/KAVYNEATC) specific for the P14 TCR (Beckmann Coulter) was used to assess TCR expression on adoptively transferred T cells. T-cell proliferation was assessed by CFSE dilution.

Intravital two-photon microscopy

DCs were pulsed with gp33–41 peptide³⁰ or OVA_{323–339}, or with medium as control, and labelled for 15 min at 37 °C with 10 μ M 5-(and 6-)-(((4-chloromethyl)benzoyl)amino)tetramethylrhodamine (CMTMR; Molecular Probes). 5 $\times$ 10⁵ DCs in 10 μ l RPMI containing 10 ng *E. coli* LPS (Sigma) were injected into the right hind footpad of recipient mice. T cells were labelled for 15 min at 37 °C with 5 μ M 5-chloromethylfluorescein diacetate (CMFDA; Molecular Probes) and (5–10) $\times$ 10⁶ cells were given to recipients by injection into the tail vein 18 h after DC injection. After 2 and 26 h, animals received 100 μ g

anti-L-selectin monoclonal antibody Mel-14 (100 µg per mouse). At various time points during the following 48 h, mice were anaesthetized by an initial intraperitoneal injection of ketamine (50 mg kg⁻¹) and xylazine (10 mg kg⁻¹). The right popliteal LN was prepared microscopically for intravital microscopy and positioned on a custom-built microscope stage. Care was taken to spare blood vessels and afferent lymph vessels. The prepared LN was submerged in normal saline and covered with a glass coverslip. A thermocouple was placed next to the LN to monitor local temperature, which was maintained at 36 °C. In some experiments the LN microcirculation was revealed by intravenous injection of 2% tetramethylrhodamine β-isothiocyanate-dextran or fluorescein isothiocyanate-dextran (500 or 2,000 kDa; Molecular Probes). Two-photon imaging was performed with an Olympus BX50WI fluorescence microscope equipped with a 20X, 0.95 numerical aperture objective (Olympus) and a Bio-Rad Radiance 2000MP Confocal/Multiphoton microscopy system, controlled by Lasersheet software (Bio-Rad). For two-photon excitation and second harmonic generation, a Tsunami Ti:sapphire laser with a 10-W MillenniaXs pump laser (Spectra-Physics) was tuned to 800 nm.

For four-dimensional analysis of cell migration, stacks of six squared *x-y* sections with 6 µm *z* spacing were acquired every 15 s with electronic zooming up to 4X to provide image volumes 30 µm in depth and 154–618 µm in width. Emitted light and second harmonic signals were detected through 400/40-nm, 525/50-nm and 620/100-nm band-pass filters with non-descanned detectors to generate three-colour images. Sequences of image stacks were transformed into volume-rendered four-dimensional movies using Volocity software (Improvision), which was also used for semi-automated tracking of cell motility in three dimensions. From *x*, *y* and *z* coordinates of cell centroids, parameters of cellular motility were calculated by using custom scripts in Matlab (MathWorks). As a measure of the cells' propensity to move away from an arbitrary point of origin, we calculated the motility coefficient by plotting each cell's absolute displacement against the square root of the time interval during which displacement occurred². Interactions between T cells and DCs were defined as physical contacts lasting more than 1 min. Only contacts whose initiation and termination was observed or that lasted for the entire observation period (60 min) were included in the analysis.

Statistical analysis

Non-normally distributed data were presented as medians and compared with the Mann-Whitney *U*-test (two groups) or the Kruskal-Wallis test followed by Dunn's test to compare multiple samples. Otherwise, data are shown as means ± s.e.m. where appropriate.

Received 15 September; accepted 18 November 2003; doi:10.1038/nature02238.

1. Stoll, S., Delon, J., Brotz, T. M. & Germain, R. N. Dynamic imaging of T cell-dendritic cell interactions in lymph nodes. *Science* **296**, 1873–1876 (2002).
2. Miller, M. J., Wei, S. H., Parker, I. & Cahalan, M. D. Two-photon imaging of lymphocyte motility and antigen response in intact lymph node. *Science* **296**, 1869–1873 (2002).
3. Bousso, P. & Robey, E. Dynamics of CD8⁺ T cell priming by dendritic cells in intact lymph nodes. *Nature Immunol.* **4**, 579–585 (2003).
4. Gowans, J. L. & Knight, E. J. The route of re-circulation of lymphocytes in the rat. *Proc. R. Soc. Lond. B* **159**, 257–282 (1964).
5. von Andrian, U. H. Immunology. T cell activation in six dimensions. *Science* **296**, 1815–1817 (2002).
6. Cahalan, M. D., Parker, I., Wei, S. H. & Miller, M. J. Two-photon tissue imaging: seeing the immune system in a fresh light. *Nature Rev. Immunol.* **2**, 872–880 (2002).
7. von Andrian, U. H. & Mempel, T. R. Homing and cellular traffic in lymph nodes. *Nature Rev. Immunol.* **3**, 867–878 (2003).
8. von Andrian, U. H. Intravital microscopy of the peripheral lymph node microcirculation in mice. *Microcirculation* **3**, 287–300 (1996).
9. Miller, M. J., Wei, S. H., Cahalan, M. D. & Parker, I. Autonomous T cell trafficking examined *in vivo* with intravital two-photon microscopy. *Proc. Natl Acad. Sci. USA* **100**, 2604–2609 (2003).
10. Gallatin, W. M., Weissman, I. L. & Butcher, E. C. A cell-surface molecule involved in organ-specific homing of lymphocytes. *Nature* **304**, 30–34 (1983).
11. Gunn, M. D. *et al.* Mice lacking expression of secondary lymphoid organ chemokine have defects in lymphocyte homing and dendritic cell localization. *J. Exp. Med.* **189**, 451–460 (1999).
12. Sallusto, F. *et al.* Rapid and coordinated switch in chemokine receptor expression during dendritic cell maturation. *Eur. J. Immunol.* **28**, 2760–2769 (1998).
13. Bajenoff, M., Granjeaud, S. & Guerdier, S. The strategy of T cell antigen-presenting cell encounter in antigen-draining lymph nodes revealed by imaging of initial T cell activation. *J. Exp. Med.* **198**, 715–724 (2003).
14. Langenkamp, A., Messi, M., Lanzavecchia, A. & Sallusto, F. Kinetics of dendritic cell activation: impact on priming of TH1, TH2 and nonpolarized T cells. *Nature Immunol.* **1**, 311–316 (2000).
15. Saxton, M. J. & Jacobson, K. Single-particle tracking: applications to membrane dynamics. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 373–399 (1997).
16. Tang, H. L. & Cyster, J. G. Chemokine up-regulation and activated T cell attraction by maturing dendritic cells. *Science* **284**, 819–822 (1999).
17. Dustin, M. L., Bromley, S. K., Zhengyan, K., Peterson, D. A. & Unanue, E. R. Antigen receptor engagement delivers a stop signal to migrating T lymphocytes. *Proc. Natl Acad. Sci. USA* **94**, 3909–3913 (1997).
18. Monks, C. R. F., Freiberg, B. A., Kupfer, H., Sciaky, N. & Kupfer, A. Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature* **394**, 82–86 (1998).
19. Grakoui, A. *et al.* The immunological synapse: a molecular machine controlling T cell activation. *Science* **285**, 221–227 (1999).
20. Lee, K. H. *et al.* T cell receptor signaling precedes immunological synapse formation. *Science* **295**, 1539–1542 (2002).
21. Gunzer, M. *et al.* Antigen presentation in extracellular matrix: interactions of T cells with dendritic cells are dynamic, short lived, and sequential. *Immunity* **13**, 323–332 (2000).
22. Ingulli, E., Mondino, A., Khoruts, A. & Jenkins, M. K. *In vivo* detection of dendritic cell antigen presentation to CD4⁺ T cells. *J. Exp. Med.* **185**, 2133–2141 (1997).

23. Norbury, C. C., Malide, D., Gibbs, J. S., Bennink, J. R. & Yewdell, J. W. Visualizing priming of virus-specific CD8⁺ T cells by infected dendritic cells *in vivo*. *Nature Immunol.* **3**, 265–271 (2002).
24. Murphy, K. M., Heimberger, A. B. & Loh, D. Y. Induction by antigen of intrathymic apoptosis of CD4⁺ CD8⁺ TCR^{lo} thymocytes *in vivo*. *Science* **250**, 1720–1723 (1990).
25. Stefanova, I., Dorfman, J. R. & Germain, R. N. Self-recognition promotes the foreign antigen sensitivity of naive T lymphocytes. *Nature* **420**, 429–434 (2002).
26. Faroudi, M., Zaru, R., Paulet, P., Muller, S. & Valitutti, S. Cutting edge: T lymphocyte activation by repeated immunological synapse formation and intermittent signaling. *J. Immunol.* **171**, 1128–1132 (2003).
27. van Stipdonk, M. J. *et al.* Dynamic programming of CD8⁺ T lymphocyte responses. *Nature Immunol.* **4**, 361–365 (2003).
28. Hurez, V. *et al.* Restricted clonal expression of IL-2 by naive T cells reflects differential dynamic interactions with dendritic cells. *J. Exp. Med.* **198**, 123–132 (2003).
29. Pircher, H., Burki, K., Lang, R., Hengartner, H. & Zinkernagel, R. M. Tolerance induction in double specific T-cell receptor transgenic mice varies with antigen. *Nature* **342**, 559–561 (1989).
30. Mora, J. R. *et al.* Selective imprinting of gut-homing T cells by Peyer's patch dendritic cells. *Nature* **424**, 88–93 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Buschman for software programming, H. Leung and B. Reinhardt for technical assistance, and D. Mathis, J. Lieberman and H. Ploegh for critical reading of the manuscript. This work was supported by grants from the NIH and the Dana Foundation Immuno-Imaging Program.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to U.H.v.A. (uva@cbr.med.harvard.edu).

Vernalization in *Arabidopsis thaliana* is mediated by the PHD finger protein VIN3

Sibum Sung & Richard M. Amasino

Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Drive, Madison, Wisconsin 53706, USA

In biennials and winter annuals, flowering is typically blocked in the first growing season. Exposure to the prolonged cold of winter, through a process called vernalization, is required to alleviate this block and permit flowering in the second growing season¹. In winter-annual types of *Arabidopsis thaliana*, a flowering repressor, *FLOWERING LOCUS C (FLC)*, is expressed at levels that inhibit flowering in the first growing season². Vernalization promotes flowering by causing a repression of *FLC* that is mitotically stable after return to warm growing conditions². Here we identify a gene with a function in the measurement of the duration of cold exposure and in the establishment of the vernalized state. We show that this silencing involves changes in the modification of histones in *FLC* chromatin.

In winter-annual types of *Arabidopsis*, flowering is delayed unless plants are vernalized. The delayed flowering is due to dominant alleles of *FRIGIDA (FRI)* and *FLC*³. *FRI* elevates expression of the MADS-box transcriptional regulator *FLC* to levels that suppress flowering^{4,5}. Vernalization promotes flowering primarily by repressing *FLC* expression^{4–7}. The repressed state of *FLC* is maintained through mitotic cell divisions after a return to warm growing conditions². Many summer-annual accessions of *Arabidopsis* flower rapidly without vernalization because such accessions lack an active *FRI* allele^{8,9} or have a weak *FLC* allele^{9,10} and thus have low levels of *FLC* expression.

To investigate the molecular mechanism of vernalization, we mutagenized a winter-annual *Arabidopsis* line (the Columbia accession into which an active *FRI* allele had been introgressed³) and

screened for mutants insensitive to vernalization. In this screen we found additional alleles of *vernalization 1* and 2 (*vrn1* and *vrn2*); these genes are necessary to maintain the repressed state of *FLC* after vernalization^{11,12}. We also identified three independent mutants that defined a new complementation group, *vernalization-insensitive 3* (*vin3-1*, *vin3-2* and *vin3-3*). In the *vin3* mutants, the vernalization response was completely blocked (Fig. 1a, b).

In *Arabidopsis*, the autonomous pathway defines a group of genes that suppress *FLC* expression¹³. Autonomous-pathway mutants in a summer-annual background exhibit a vernalization-responsive late-flowering phenotype similar to winter annuals because both have a greater expression of *FLC* before vernalization^{6,7,13}. The *vin3* lesion blocks the vernalization response in the autonomous-pathway mutant *luminidependens* (Fig. 1c). Vernalization also promotes flowering independently of *FLC* repression; specifically, vernalization can cause more rapid flowering in short days in both the wild type and a *flc*-null mutant⁷. The promotion of flowering by vernalization in short days is prevented in a *vin3 flc* double mutant (Fig. 1d). Thus, *VIN3* is required for all situations, both *FLC*-dependent and *FLC*-independent, in which vernalization can promote flowering. In both the parental winter-annual *FRI*-containing line and in the summer-annual *Col* background (*fri*), the *vin3* mutant displayed no pleiotropic phenotypes and had no effect on flowering other than the block to vernalization (Fig. 1a, b, and data not shown).

In *vin3* mutants, a vernalizing cold treatment is not able to repress *FLC* expression during and after cold exposure (Fig. 1e). This is in contrast to the behaviour of *vrn1* and *vrn2* mutants in which vernalization-mediated *FLC* repression is observed during the

period of cold exposure; in other words, *vrn1* and *vrn2* behave like the wild type with regard to *FLC* repression during a vernalizing cold treatment, but this repression is not maintained in warm growing conditions (Fig. 2b, c, and refs 11 and 12). Thus, unlike *VRN1* and *VRN2*, which are involved in the maintenance of *FLC* repression, *VIN3* has a function in the establishment of *FLC* repression during vernalization.

We cloned *VIN3* by a map-based approach (a rescued line is shown in Fig. 1a; *VIN3* is At5g57380; see www.arabidopsis.org). *VIN3* is composed of four exons encoding a 600-amino-acid protein that contains a putative nuclear localization signal and two domains found in other proteins (Fig. 2a): a PHD (plant homeodomain) domain, which is often found in proteins in chromatin remodelling complexes¹⁴, and a FNIII (fibronectin type III) domain, which is often involved in protein-protein interactions¹⁵.

A requirement for vernalization is an adaptation for ensuring that flowering does not occur before the onset of winter. To prevent flowering during short-term temperature fluctuations in the autumn, the vernalization process has evolved to distinguish a long exposure to cold from shorter cold exposures². The pattern of *VIN3* induction is consistent with a role for *VIN3* in the process of distinguishing the duration of cold exposure (Fig. 2b). *VIN3* is expressed only after a duration of cold exposure that is effective for vernalization, and *FLC* repression does not occur until *VIN3* is induced (Fig. 2b). Moreover, the level of *VIN3* expression is correlated with both the duration of cold and the degree of *FLC* repression; after a long but not vernalization-saturating period of cold exposure, *VIN3* is partly induced and *FLC* is partly repressed (Fig. 2b; compare expression levels at 20 and 40 days of vernaliza-

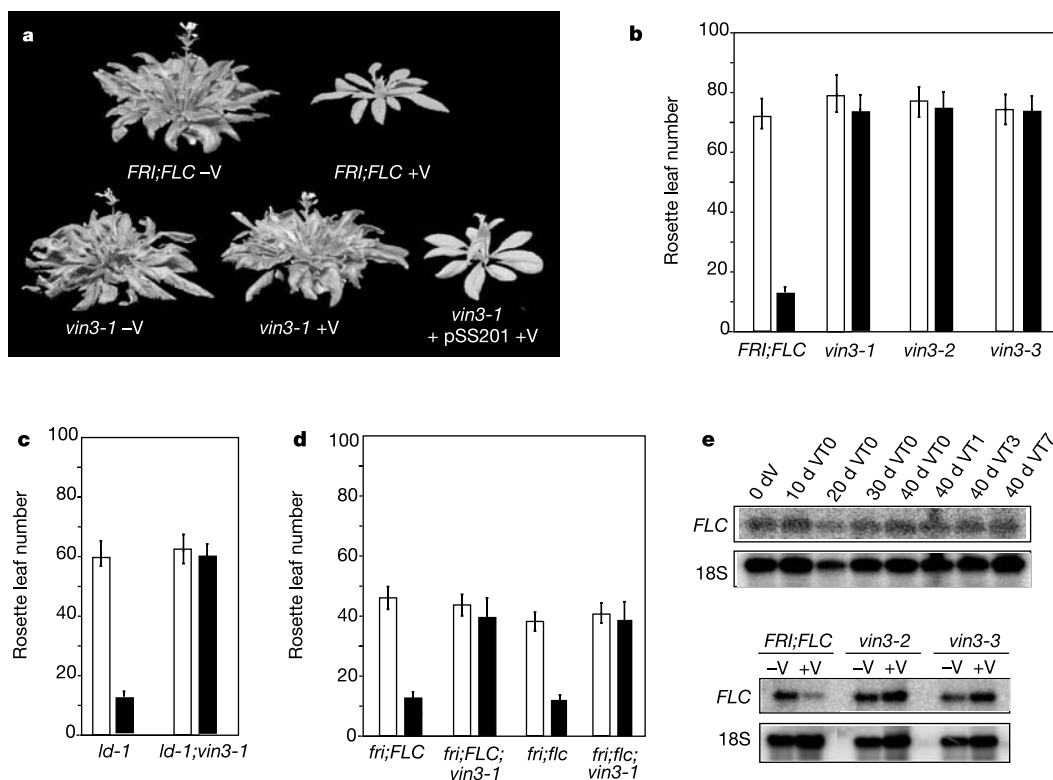


Figure 1 Characterization of *vernalization-insensitive 3* mutants. **a**, *vin3* blocks the vernalization response. Top, parental winter-annual line with and without vernalization. Bottom, *vin3-1* mutant with (+V) and without (-V) vernalization and rescue of *vin3-1* with a *VIN3* transgene. **b**, Block of the vernalization response in *vin3* mutants. Open bars, leaf number at flowering of non-vernalized plants; filled bars, values for plants vernalized for 40 days as described⁶. **c**, *vin3* causes vernalization insensitivity in the *luminidependens*

mutant. **d**, *vin3* blocks the vernalization response in a *flc* mutant in short days. **e**, *FLC* is not repressed by vernalization in *vin3* mutants. Top panel, length of exposure to 2–4 °C is shown as 'd V'. T0 samples were harvested directly from cold-grown plants. T1, T3 and T7 indicate the number of days for which the plants were grown at 22 °C after 40 days of cold. Bottom panel: +V samples were vernalized for 40 days, followed by 7 days at 22 °C.

tion). In addition, *VIN3* mRNA becomes undetectable within 3 days after return to a warm growth temperature (Fig. 2b, c). Thus, in contrast to *VRN1* and *VRN2*, which are constitutively expressed^{11,12}, *VIN3* is expressed only during cold exposure. Furthermore, the appearance of *VIN3* expression is a marker for cold exposures of a duration that are effective for vernalization. *VIN3* expression is induced by cold exposure in *vrn1* (Fig. 2c) and *vrn2* (data not shown), indicating that *VIN3* is upstream of *VRN1* and *VRN2*.

Vernalization causes an epigenetic change that provides competence to flower only in specific tissues such as shoot and root apices^{2,16,17}. The pattern of *VIN3* expression, determined by a *VIN3* promoter fusion to β -Glucuronidase (*GUS*), matches the regions in which vernalization is effective (Fig. 2d). Furthermore, *VIN3* has the same pattern of expression as its target *FLC* (Fig. 2d). The *FLC-GUS* construct is subject to repression by vernalization in a wild-type background, whereas in a *vin3* mutant *FLC-GUS* repression does not occur during or after vernalization (Fig. 2e). In contrast, the *FLC-GUS* construct is repressed in a *vrn1* mutant during vernalization (Fig. 2e). The *VIN3-GUS* construct is expressed only when the plants are exposed to a long period of cold, and expression disappears after the plants are grown in warm conditions (Fig. 2d). Thus, the cold and tissue-specific expression of *VIN3* is likely to be controlled at the transcriptional level. These results, together with the RNA and phenotypic data described above, further support a role for *VIN3* in the vernalization-mediated

establishment of *FLC* repression in the shoot and root apex.

The epigenetic nature of *FLC* repression by vernalization, and the involvement of proteins related to chromatin-modelling proteins in this process, prompted us to examine whether histone modifications were involved in vernalization-mediated silencing of *FLC*. In humans and *Drosophila*, Su(z)12 polycomb group proteins, which are homologues of *VRN2*, mediate gene silencing by a series of histone modifications leading to methylation of histone H3 at Lys 9 (ref. 18). Methylation of Lys 9 is a marker for heterochromatin formation, and H3 Lys 9 methylation is thought to recruit heterochromatin protein 1 (and possibly to facilitate DNA methylation) to ensure a mitotically repressed state¹⁹. During the silencing process, a transiently acting repressor is often required to target specific genes for histone modifications¹⁸. One of the initial events that requires a transiently acting repressor can be histone deacetylation²⁰.

We therefore examined the acetylation of histone H3 at Lys 9 and Lys 14 in chromatin of *FLC*. We observed that a region of intron I (V1) and a region upstream of the transcriptional start site (P1) of *FLC* showed decreased acetylation during vernalization, and after transfer to warm growing conditions this decrease was maintained (Fig. 3a). However, *FLC* chromatin in the region (U1) encoding the 3' untranslated region showed little change in acetylation levels. In *vin3* the vernalization-mediated changes in *FLC* acetylation do not occur (Fig. 4a; data shown for the V1 region). In *vrn1* and *vrn2* a decreased acetylation of *FLC* occurs during vernalization, but this

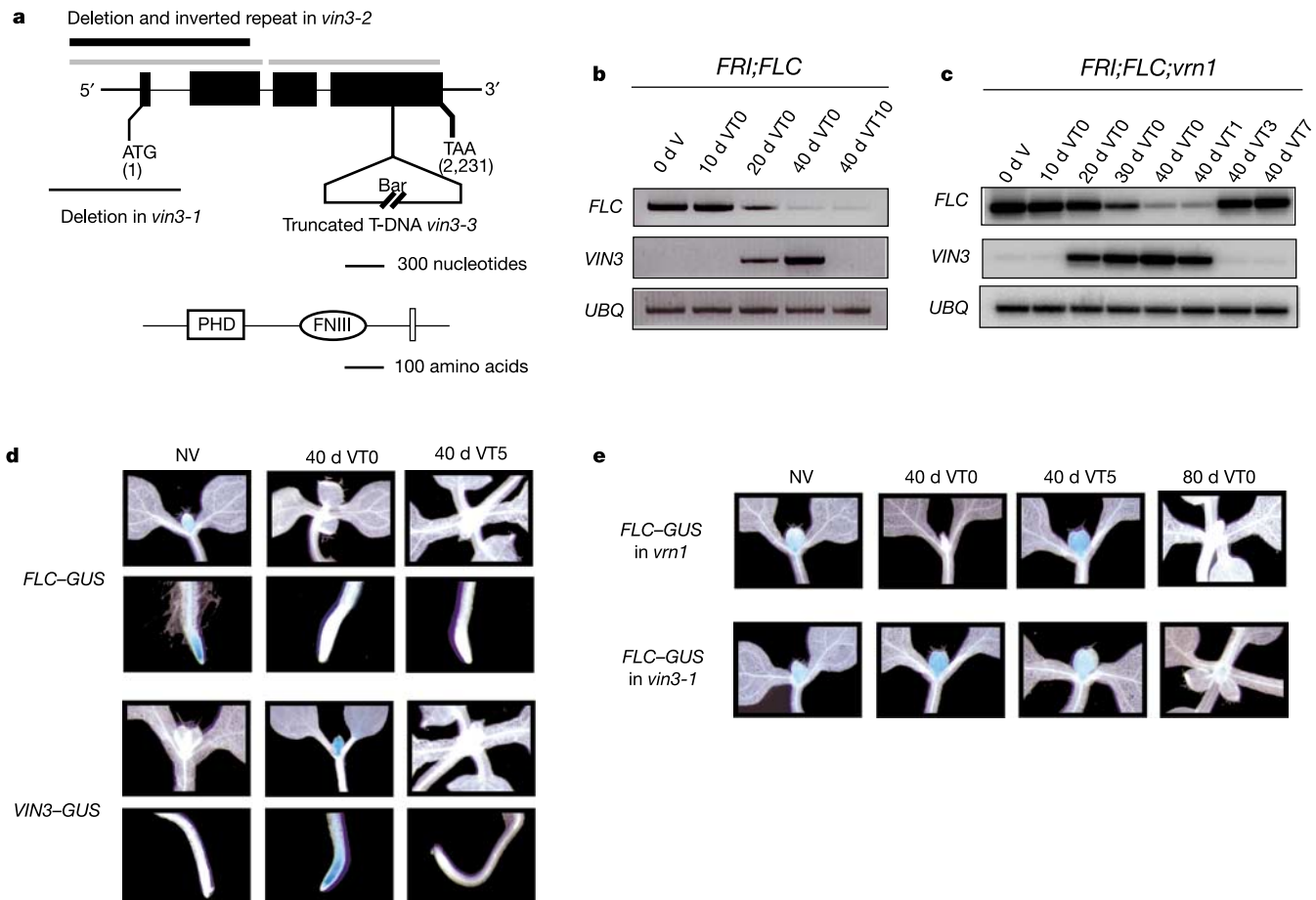


Figure 2 *VIN3* cloning and expression pattern. **a**, Positions of the domains in *VIN3* (PHD finger, fibronectin type III, and (small rectangle) nuclear localization signal) and the lesions in *vin3* alleles. *vin3-1* is a fast-neutron allele, and *vin3-2* and *vin3-3* are T-DNA insertion alleles. **b**, RT-PCR analyses of the effect of vernalization on *VIN3* and *FLC* expression in the winter-annual wild type. Samples are as described in Fig. 1e. **c**, RT-PCR

analyses of *VIN3* and *FLC* expression in *vrn1*. **d**, Spatial pattern of *VIN3* and *FLC* expression determined by reporter gene (*GUS*) fusions in the wild type without vernalization (NV), after 40 days of vernalization (40 d VT0), and after vernalization and 5 days at 22 °C (40 d VT5). **e**, *FLC* expression in *vrn1* and *vrn3*. Designations are as in **d**.

decrease is not maintained in warm growing conditions (Fig. 4a, b). The lack of changes in *FLC* histone acetylation patterns in *vin3* mutants and the transient cold-specific expression of *VIN3* both in the wild type and in *vrn1* and *vrn2* during vernalization indicates that *VIN3* participates in a transient repression of *FLC* that involves histone deacetylation, and that this *VIN3*-mediated process is required for the establishment of *FLC* silencing.

A candidate modification for the maintenance of vernalization-mediated *FLC* repression is H3 dimethylation at Lys 9 (ref. 19). Indeed, we find that in the wild type, dimethylation levels at Lys 9 of H3 in chromatin of the *FLC* promoter and intron regions (P1 and V1) increase during vernalization and that this increase is main-

tained after vernalization (Fig. 3b). In *vin3*, as well as in *vrn1* and *vrn2*, the vernalization-mediated increase in H3 dimethylation at Lys 9 does not occur (Fig. 4). We also observe that in the wild type the level of dimethylation at Lys 27 of H3 increases during and after vernalization (Fig. 3b). Interestingly, the increase in H3 dimethylation at Lys 27 occurs in *vrn1* but not in *vrn2* (Fig. 4b, c) indicating that *VRN1* is not required for H3 dimethylation at Lys27 and that this dimethylation alone is not sufficient to maintain *FLC* repression. The lack of H3 dimethylation at both Lys 9 and Lys 27 in *vrn2* indicates that *VRN2* activity is required for both dimethylations, whereas *VRN1* activity is necessary for H3 dimethylation at Lys 9. *VRN2* homologues in *Drosophila* and human (*SUPPRESSOR OF ZESTE*; *Su(z)12*) are involved in H3 methylation¹⁸, indicating that the role of this type of polycomb group protein in gene silencing is evolutionarily conserved.

To address whether *VIN3* can interact directly with the *FLC* locus, we evaluated whether *VIN3* antiserum could precipitate *FLC* chromatin (Fig. 3c). Precipitation of chromatin from vernalized plants with *VIN3* antiserum greatly enriched for *FLC* compared

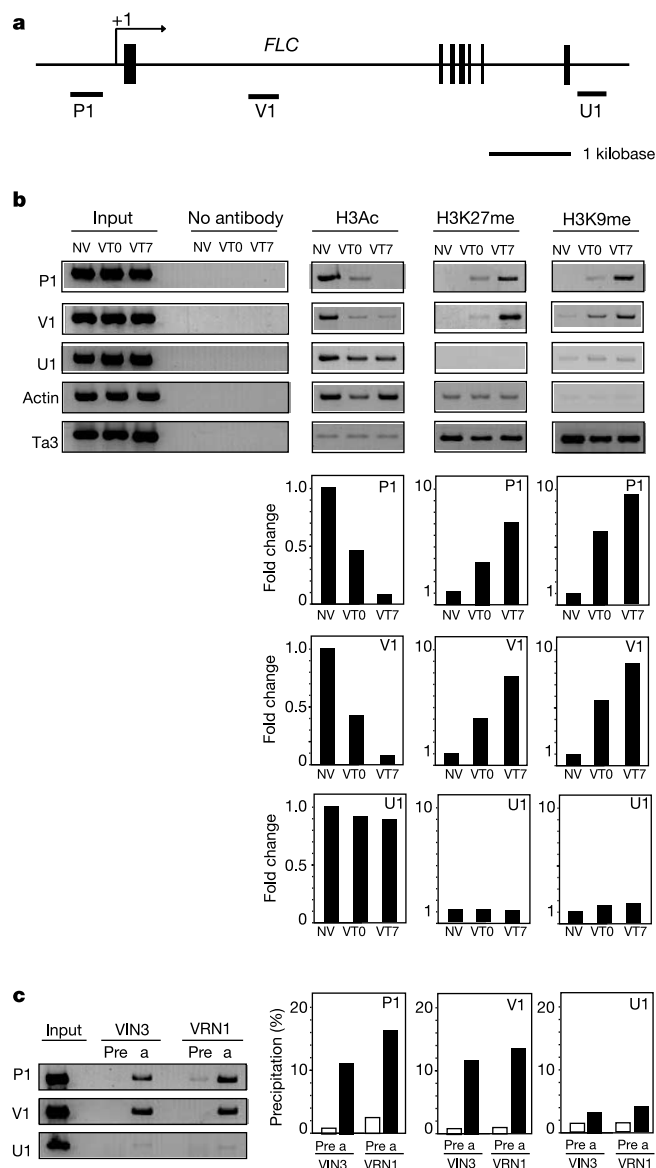


Figure 3 ChIP assays in wild-type winter-annual *Arabidopsis*. **a**, Regions of the *FLC* locus examined (see Methods for details). Coding regions are indicated by vertical lines. **b**, ChIP with antibodies against modified histones. Samples are identified as follows: NV, non-vernalized; VT0, vernalized for 40 days; VT7, vernalized for 40 days and subsequently grown for 7 days at 22°C. Sample preparation was as described previously³⁰. Signal intensities were normalized relative to actin and Ta3 control reactions with ImageQuant, and fold changes are presented in the bar graphs. **c**, ChIP with *VIN3* and *VRN1* antisera. Samples were from plants vernalized for 40 days without exposure to warm temperature. Signal intensities were normalized relative to input signal intensities. Pre, pre-immunized serum.

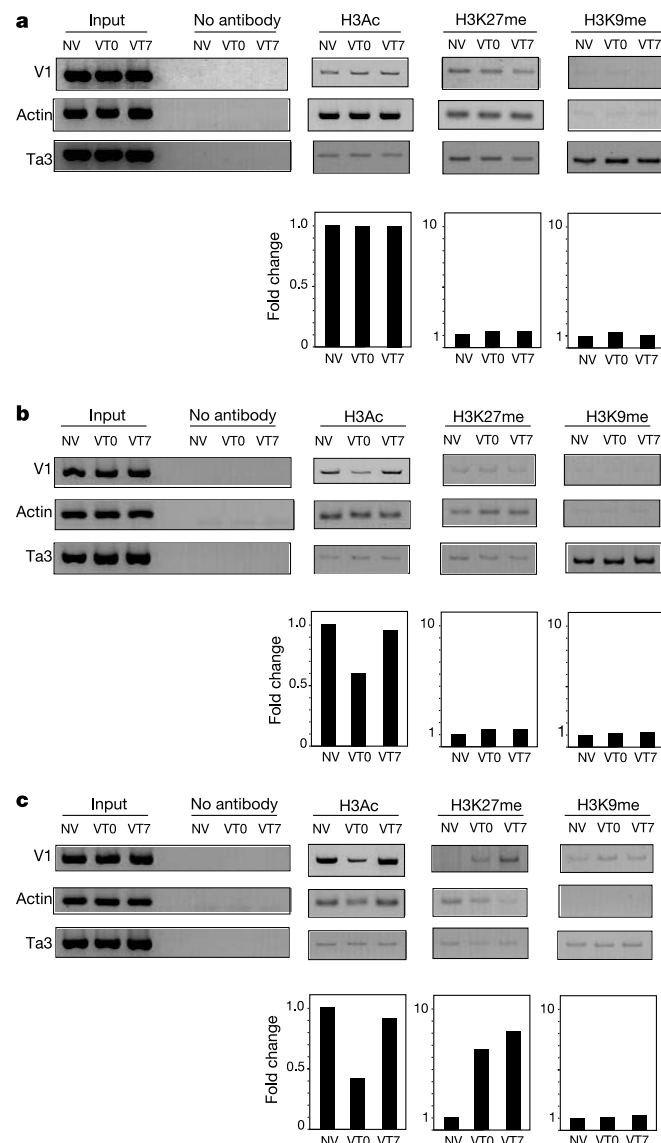


Figure 4 Chromatin immunoprecipitation (ChIP) assays in vernalization mutants. ChIP assays and designations are as described in Fig. 3. Genotypes were *vin3* (**a**), *vrn2* (**b**) and *vrn1* (**c**).

with the preimmune control. In addition, VRN1 antiserum greatly enriched for *FLC* chromatin compared with the preimmune control. The regions of the promoter (P1) and the intron (V1) were enriched in these immunoprecipitations to a much greater extent than the 3' end (U1) of *FLC*. Thus, the regions of *FLC* chromatin with which VIN3 and VRN1 interact in immunoprecipitation experiments correspond to the regions of *FLC* chromatin in which changes in histone modification occur. The interaction of VIN3 and VRN1 with the 5' promoter and the first intron regions and the vernalization-mediated histone modifications of these regions are consistent with the observation that the regions are required for a proper vernalization response²¹ and that a DNase I-hypersensitive site in the intron region becomes insensitive after vernalization¹¹.

The epigenetic nature of vernalization was first recognized as a memory of winter in which the competence of an apical meristem to initiate flowering persists through mitotic cell divisions^{22,23}. Our results provide a framework for the molecular basis of vernalization in *Arabidopsis*. The establishment of *FLC* silencing (and *FLC*-independent aspects of the vernalized state) requires the induction of VIN3 expression by exposure to prolonged cold. VIN3 expression is necessary for deacetylation of *FLC*, which in turn leads to histone methylation and the formation of mitotically stable heterochromatin at the *FLC* locus by a process involving VRN1 and VRN2. VIN3 is induced by cold only in tissues (root and shoot apices) in which vernalization occurs, and the spatial restriction of VIN3 is likely to contribute to the restriction of the vernalization process to these tissues.

VRN1 and VRN2 are expressed constitutively^{11,12}, and the recruitment of VRN1 and VRN2 to target genes such as *FLC* during vernalization is likely to require a factor to provide specificity. The spatial expression pattern and cold-specific induction of VIN3 and the interaction of VIN3 with *FLC* chromatin indicates that VIN3 could be part of a transiently acting complex that leads to the initial establishment of target gene repression during vernalization. It is consistent with this model that VIN3 is clearly the most upstream component of vernalization so far identified because *vin3* is the only mutant in which *FLC* repression is blocked during cold exposure, and in which no histone modifications are detected during vernalization. Moreover, VIN3 is induced in *vrn1* and *vrn2*, and VIN3-mediated *FLC* silencing during cold exposure occurs in *vrn1* and *vrn2*.

The additional components that interact with VIN3, and VRN1 and VRN2, to repress *FLC* during and after vernalization are not known. One component might be a protein similar to ENHANCER OF ZESTE (E(z)). In human and *Drosophila*, E(z) acts as a histone methyltransferase in a complex with the VRN2 homologue Su(z)12, and this complex is involved in polycomb-complex-mediated gene repression¹⁸. MEDEA, CURLY LEAF and EZA1 are the closest relatives of E(z) in *Arabidopsis*²⁴. Lesions in CURLY LEAF cause ectopic expression of the floral homeotic gene AGAMOUS²⁴ and do not affect the vernalization process (not shown). It is possible that MEDEA or EZA1 has a function in vernalization, or that there is redundancy between E(z) relatives in this process, or that other proteins are involved in vernalization-mediated histone methylation. It will be interesting to address nature of the complex that leads to establishment of *FLC* repression during vernalization.

One of the remarkable aspects of the vernalization process is that it has evolved to sense the prolonged cold of winter. Exposure to short periods of cold that are sufficient to initiate other cold responses such as cold acclimation²⁵ do not initiate a vernalization response². Part of the mechanism for measuring the period of cold exposure during vernalization is the requirement for several weeks of cold treatment for VIN3 induction. However, constitutive VIN3 expression is not sufficient to substitute for cold treatment (data not shown); thus, VIN3 must interface with other cold-induced changes to cause vernalization. The recent demonstration that the PHD

finger domain can bind phosphoinositides²⁶ raises the possibility that VIN3 could be a nuclear phosphoinositide receptor, and that perhaps VIN3 is involved in perception of changes in the spectrum of nuclear phosphoinositides that might occur during cold exposure. It will be interesting to further explore the mechanism by which prolonged cold induces VIN3 expression and achieves the vernalized state. □

Methods

Plant materials

The parental *FRI-Sf2* in Columbia was described previously³. pSKI015 T-DNA²⁷ was used for T-DNA insertional mutagenesis of this line, and fast-neutron mutagenesis was as described⁴.

Mapping

A Landsberg *erecta* line into which was introgressed *FRI-Sf2* and *FLC-Col* (ref. 3) was crossed to *vin3-1*, *vin3-2*, *vin3-3* which are in the *FRI-Sf2* in Columbia background to create mapping populations; 2,345 individual F₂ plants were used to localize VIN3.

RNA analyses were performed as described for blotting²⁸ and reverse transcriptase polymerase chain reaction (RT-PCR)²⁹.

Chromatin immunoprecipitation (ChIP) assays

Formaldehyde cross-linked plant materials were used for ChIP assays as described³⁰. Anti-acetyl-histone H3, anti-dimethyl-histone H3 [Lys 9] and anti-dimethyl-histone H3 [Lys 27] antibodies from Upstate Biotechnology (Waltham, Massachusetts, USA) were used for precipitation. Rabbit polyclonal antibodies against VRN1 and VIN3 were prepared by Covance (Denver, Pennsylvania, USA) against synthetic peptides SSQGNVCVYPETTS (VIN3) and AFSVYIFNLHSEIN (VRN1) and were conjugated to keyhole limpet haemocyanin. Three pairs of primers that amplified the following regions of the *FLC* locus were used to assess immunoprecipitation enrichment by PCR. The regions of the primer sets (P1, V1 and U1), relative to the start of transcription, were P1 (−495 to −470 and −222 to −197), V1 (+1436 to +1461 and +1699 to +1724) and U1 (+5805 to +5830 and +6019 to +6044).

Received 22 August; accepted 27 October 2003; doi:10.1038/nature02195.

1. Chouard, P. Vernalization and its relations to dormancy. *Annu. Rev. Plant Physiol.* **11**, 191–238 (1960).
2. Michaels, S. & Amasino, R. Memories of winter: vernalization and the competence to flower. *Plant Cell Environ.* **23**, 1145–1154 (2000).
3. Lee, I., Michaels, S. D., Masshardt, A. S. & Amasino, R. M. The late-flowering phenotype of *FRIGIDA* and *LUMINIDEPENDENS* is suppressed in the Landsberg *erecta* strain of *Arabidopsis*. *Plant J.* **6**, 903–909 (1994).
4. Michaels, S. & Amasino, R. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949–956 (1999).
5. Sheldon, C. C. *et al.* The *FLF* MADS box gene. A repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* **11**, 445–458 (1999).
6. Sheldon, C. C., Rouse, D. T., Finnegan, E. J., Peacock, W. J. & Dennis, E. S. The molecular basis of vernalization: the central role of *FLOWERING LOCUS C* (*FLC*). *Proc. Natl Acad. Sci. USA* **97**, 3753–3758 (2000).
7. Michaels, S. D. & Amasino, R. M. Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. *Plant Cell* **13**, 935–942 (2001).
8. Johanson, U. *et al.* Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science* **290**, 344–347 (2000).
9. Gazzani, S., Gendall, A. R., Lister, C. & Dean, C. Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol.* **132**, 1107–1114 (2003).
10. Michaels, S. D., He, Y., Scortecci, K. C. & Amasino, R. M. Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of summer-annual flowering behavior in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **100**, 10102–10107 (2003).
11. Gendall, A. R., Levy, Y. Y., Wilson, A. & Dean, C. The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. *Cell* **107**, 525–535 (2001).
12. Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R. & Dean, C. Multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control. *Science* **297**, 243–246 (2002).
13. Simpson, G. G. & Dean, C. *Arabidopsis*, the Rosetta stone of flowering time? *Science* **296**, 285–289 (2002).
14. Fair, K. *et al.* Protein interactions of the MLL PHD fingers modulate MLL target gene regulation in human cells. *Mol. Cell Biol.* **21**, 3589–3597 (2001).
15. Main, A. L., Harvey, T. S., Baron, M., Boyd, J. & Campbell, I. D. The three-dimensional structure of the tenth type III module of fibronectin: an insight into RGD-mediated interactions. *Cell* **71**, 671–678 (1992).
16. Wellensiek, S. J. Dividing cells as the locus for vernalization. *Nature* **195**, 307–308 (1962).
17. Burn, J. E., Bagnall, D. J., Metzger, J. D., Dennis, E. S. & Peacock, W. J. DNA methylation, vernalization, and the initiation of flowering. *Proc. Natl Acad. Sci. USA* **90**, 287–291 (1993).
18. Kuzmichev, A. R. D., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of Zeste protein. *Genes Dev.* **16**, 2893–2905 (2002).
19. Richards, E. J. & Elgin, S. C. Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell* **108**, 489–500 (2002).
20. Kehle, J. *et al.* dMi-2, a hunchback-interacting protein that functions in polycomb repression. *Science* **282**, 1897–1900 (1998).
21. Sheldon, C. C., Conn, A. B., Dennis, E. S. & Peacock, W. J. Different regulatory regions are required for

- the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. *Plant Cell* **14**, 2527–2537 (2002).
22. Lang, A. & Melchers, G. Vernalisation und Devernalisation bei einer zweijährigen Pflanze. *Z. Naturforsch* **2b**, 444–449 (1947).
 23. Lang, A. in *Encyclopedia of Plant Physiology* Vol. 15 Part 1 (ed. Ruhland, W.) 1371–1536 (Springer, Berlin, 1965).
 24. Reyes, J. C., Hennig, L. & Gruissem, W. Chromatin-remodeling and memory factors. New regulators of plant development. *Plant Physiol.* **130**, 1090–1101 (2002).
 25. Thomashow, M. F. So what's new in the field of plant cold acclimation? Lots! *Plant Physiol.* **125**, 89–93 (2001).
 26. Gozani, O. et al. The PHD finger of the chromatin-associated protein ING2 functions as a nuclear phosphoinositide receptor. *Cell* **114**, 99–111 (2003).
 27. Weigel, D. et al. Activation tagging in *Arabidopsis*. *Plant Physiol.* **122**, 1003–1013 (2000).
 28. Michaels, S. D. & Amasino, R. M. The gibberellic acid biosynthesis mutant *gal-3* of *Arabidopsis thaliana* is responsive to vernalization. *Dev. Genet.* **25**, 194–198 (1999).
 29. Michaels, S. D. et al. *AGL24* acts as a promoter of flowering in *Arabidopsis* and is positively regulated by vernalization. *Plant J.* **33**, 867–874 (2003).
 30. Johnson, L., Cao, X. & Jacobsen, S. Interplay between two epigenetic marks. DNA methylation and histone H3 lysine 9 methylation. *Curr. Biol.* **12**, 1360–1367 (2002).

Acknowledgements We thank S. Michaels for creating and helping to screen the mutant populations that provided the *vin3* mutant, for the FLC–GUS construct, and for advice and insight; and M. Doyle for help in the preparation of the manuscript. R.M.A. thanks C. O. Miller and J. A. D. Zeevaert for encouragement and mentoring in flowering-time regulation. This work was supported by the College of Agricultural and Life Sciences and the Graduate School of the University of Wisconsin, by the United States Department of Agriculture National Research Initiative Competitive Grants Program, and by a grant to R.M.A. from the National Science Foundation. The creation of insertion mutant lines was supported by a National Science Foundation grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.M.A. (amasino@biochem.wisc.edu). The GenBank accession number for the VIN3 cDNA is AY446960.

Vernalization requires epigenetic silencing of *FLC* by histone methylation

Ruth Bastow¹, Joshua S. Mylne¹, Clare Lister¹, Zachary Lippman², Robert A. Martienssen² & Caroline Dean¹

¹Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Colney, Norwich NR4 7UH, UK

²Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA

To ensure flowering in favourable conditions, many plants flower only after an extended period of cold, namely winter. In *Arabidopsis*, the acceleration of flowering by prolonged cold, a process called vernalization, involves downregulation of the protein FLC, which would otherwise prevent flowering^{1,2}. This lowered FLC expression is maintained through subsequent development by the activity of *VERNALIZATION* (*VRN*) genes^{3,4}. *VRN1* encodes a DNA-binding protein⁴ whereas *VRN2* encodes a homologue of one of the Polycomb group proteins, which maintain the silencing of genes during animal development³. Here we show that vernalization causes changes in histone methylation in discrete domains within the *FLC* locus, increasing dimethylation of lysines 9 and 27 on histone H3. Such modifications identify silenced chromatin states in *Drosophila* and human cells^{5–7}. Dimethylation of H3 K27 was lost only in *vrn2* mutants, but dimethylation of H3 K9 was absent from both *vrn1* and *vrn2*, consistent with *VRN1* functioning downstream of *VRN2*. The epigenetic memory of winter is thus mediated by a 'histone code' that specifies a silent chromatin state conserved between animals and plants.

The requirement for vernalization is an important trait in crop

breeding and has resulted in the availability of winter- and spring-sown varieties, which has significantly extended the geographical range for the farming of many crops. An understanding of vernalization in the control of flowering has emerged from a molecular genetic analysis in *Arabidopsis*⁸. The pathways conferring a requirement for and ability to respond to vernalization converge on the regulation of *FLC*^{1,2}. FLC is a MADS box transcriptional regulator that functions as a floral repressor by inhibiting the activation of a set of genes required for transition of the apical meristem to a reproductive fate^{1,2,9–13}. *FRIGIDA* (*FRI*) confers a vernalization requirement by upregulating *FLC* expression, and this is antagonized by vernalization, which reduces *FLC* levels. Genes classified in the autonomous floral pathway, such as *FCA*, function in parallel with vernalization to downregulate *FLC*. *fca* mutants are late-flowering owing to increased *FLC*, and this phenotype is suppressed by vernalization^{1,2}. Once *FLC* transcript levels are downregulated by prolonged cold exposure, they remain low during subsequent development, and this mitotic stability suggests that vernalization has an epigenetic basis. This led to the idea that DNA methylation might have a role in *FLC* regulation¹⁴. Despite the observation that *FLC* levels were reduced in plants expressing an antisense DNA methyltransferase¹, bisulphite sequencing has shown that there is no change in *FLC* DNA methylation caused by vernalization (J. Finnegan, personal communication).

To investigate the molecular basis of the cold-induced repression of *FLC*, a series of *Arabidopsis vrn* mutants, defective in the acceleration of flowering by vernalization, were identified¹⁵. *VRN1* and *VRN2* were shown to be required for the maintenance of *FLC* repression during subsequent development following prolonged cold exposure^{3,4}. *VRN2* is a zinc-finger protein homologous to SU(Z)12, a member of the ESC-E(Z) Polycomb group (Pc-G) complex, which maintains the silencing of *Drosophila HOX* genes¹⁶. Pc-G proteins act in multiprotein complexes to maintain a silenced chromatin state by modifying specific amino acids in the amino-terminal tails of histones through deacetylation or methylation¹⁷. The ESC-E(Z) Pc-G complex in mammals and flies has been shown to contain histone lysine methyltransferase activity for K27 and possibly K9 of histone H3^{5,6,18,19}. Methylated K9 of histone H3 triggers the binding of HETEROCHROMATIN PROTEIN 1, leading to heterochromatin formation and silencing^{7,20,21}. Phosphorylation, ubiquitination and sumoylation of specific amino acids of histone tails have also been described¹⁷. The varying marks placed on the histones combine to form a 'histone code' that specifies a chromatin state and subsequently determines whether a locus is transcriptionally active or silent²².

To investigate whether histone modifications were involved in the vernalization-dependent regulation of *FLC*, we analysed the chromatin environment of the *FLC* locus by chromatin immunoprecipitation (ChIP). In this technique, chromatin is isolated from cells, sheared and immunoprecipitated with a range of antibodies specific to different histone modifications. The immunoprecipitates are then examined for specific DNA sequences by polymerase chain reaction (PCR) analysis. ChIP analysis on *fca-1* plants, undertaken in at least three separate experiments, using antibodies specific to H3 dimethyl K9, H3 dimethyl K27 and H3 dimethyl K4 (a modification associated with active loci²³) revealed vernalization-specific modifications at *FLC* (Fig. 1). Sequences from the 5' region of *FLC*, regions shown as A and B in Fig. 1, which cover the *FLC* promoter and the first exon, were found to show higher levels of H3 K9 dimethylation in vernalized tissue. H3 K27 dimethylation was higher in vernalized tissue, predominantly in region B, covering the 5' end of the transcript, but was also found in some experiments in the promoter region (region A; data not shown). H3 K4 dimethylation was associated with all regions of *FLC* examined in non-vernalized tissue, but was reduced in region E, in the middle of intron 1, in vernalized samples. This could result from the targeted loss of H3 dimethyl K4 nucleosomes from this region or possibly

from the cross-linking of vernalization-specific proteins that block the epitope. A Mutator-like transposon has been found at the 3' end of intron 1 of the Landsberg *erecta* (*Ler*) *FLC* allele, and this seems to restrict high levels of *FLC* RNA accumulation^{24,25}. Consistent with this, chromatin containing this transposon showed vernalization-independent H3 K9 dimethylation (data not shown). However, the transposon does not affect the vernalization response or the maintenance of *FLC* repression, so is unlikely to influence the vernalization-induced chromatin changes at *FLC*. It was interesting to note that H3 K4 methylation was still found at most regions

within the *FLC* locus after vernalization. This could mean that K4 dimethylation is not reduced in silenced *FLC*²³, perhaps reflecting the reversibility of *FLC* repression later in development. Alternatively, it may reflect the fact that *FLC* expression is not silenced by vernalization in all cells. We have evidence for the latter from analysis of a *FLC*–*GUS* (β -glucuronidase) translational fusion in a transgenic *fca-1* line. Vernalization reduced *FLC*–*GUS* expression in most cells, but not in those of the vasculature, with strongest expression remaining in the vasculature of the cotyledons and hypocotyl (Fig. 2). In some of the ChIP experiments, an antibody to acetylated H3 was included (Figs 1 and 3); however, changes in *FLC* histone acetylation as significant as those seen for histone methylation were not observed, so this was not pursued further.

In summary, the ChIP analysis showed that regions within the promoter, the 5' end of the transcript and intron 1 of *FLC* undergo vernalization-specific histone methylation changes that are normally associated with the maintenance of transcriptional repression. These regions can be defined only approximately (within ~750 base pairs (bp)) as sonication sheared the chromatin to between 500 and 1,500 bp. The non-uniform distribution of the histone modifications observed at *FLC* is unlike the situation at heterochromatic transposons from the knob on chromosome 4S, which are more uniformly associated with H3 dimethyl K9 (ref. 26, and Z.L. and R.A.M., unpublished data). In this respect, these modified domains within the *FLC* locus resemble the Polycomb response elements in *Drosophila*, *cis*-acting elements that recruit Pc-G complexes. Whether they function in this way remains to be established.

Analysis of deletion derivatives of a *FLC*–*GUS* fusion has defined *cis*-acting domains important for *FLC* regulation²⁷, and these coincide with the regions showing vernalization-induced histone methylation changes. Region E, which is less abundant in the H3 dimethyl K4 antibody immunoprecipitate after vernalization, is contained within a region defined as required for the maintenance of *FLC* repression. Intron 1 seems to contain multiple maintenance elements, as deletion of either end of the intron did not disrupt the maintenance of *FLC* repression. Region A, which shows H3 dimethyl K9 (and sometimes dimethyl K27) modification after vernalization, corresponds almost exactly to a potential negative regulatory promoter element whose deletion leads to higher *FLC* expression. Region B, which showed both H3 dimethyl K27 and H3 dimethyl K9 modification after vernalization, is part of a much larger region defined as being required for *FLC* expression in non-vernalized plants and for repression by vernalization, so a role in the maintenance of repression could not be directly tested²⁷. Further definition of *FLC cis*-acting domains *in vivo* and further definition of the domains showing histone modifications will be important to

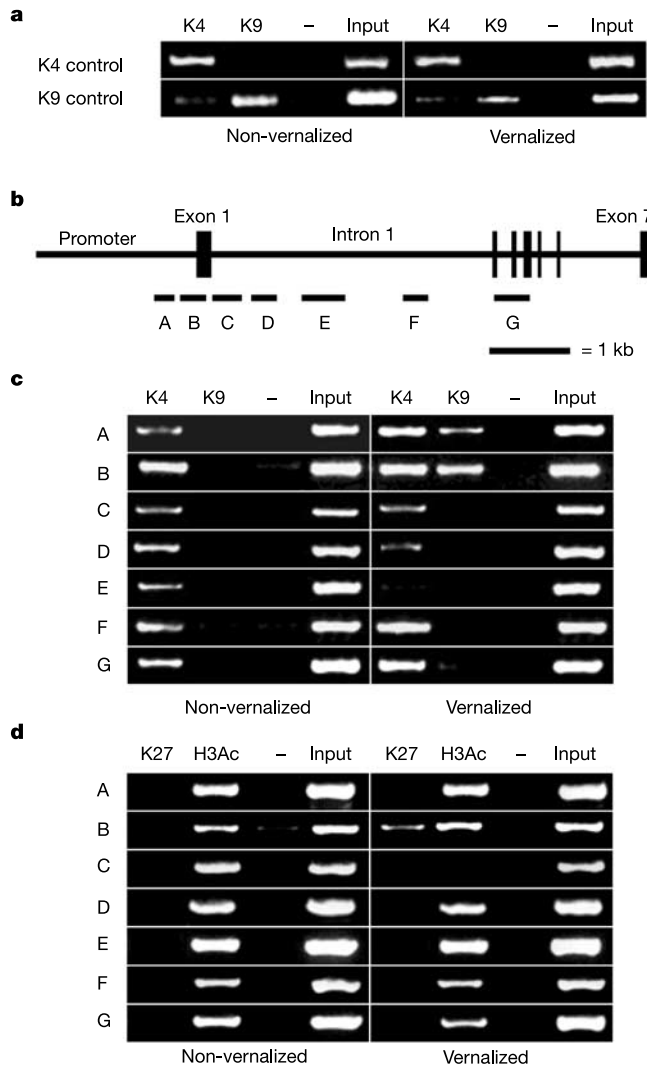


Figure 1 Histone modifications at *FLC* associated with vernalization. **a**, ChIP antibody controls: K4 primers amplify the transcribed phosphofructokinase gene on BAC T24H24.15, K9 control primers amplify the silenced Cinfu-like retrotransposon on BAC T5L23.29. **b**, Genomic structure of *Arabidopsis FLC* (Columbia accession) and regions examined by ChIP. Black boxes represent exons; lines represent promoter and introns. Regions amplified by PCR are shown below as bars labelled A to G. For primer sequences, see Supplementary Fig. 3. **c**, ChIP analysis using H3 dimethyl K4 and H3 dimethyl K9 antibodies. *fca-1* plants were grown with and without prior vernalization. Vernalization resulted in increased precipitation of regions A and B with the H3 dimethyl K9 antibody, and reduced precipitation of region E with the H3 dimethyl K4 antibody. **d**, ChIP analysis using antibodies against H3 dimethyl K27 and acetylated H3 with *fca-1* plants. Region B was immunoprecipitated with the H3 dimethyl K27 antibody in vernalized seedlings. Identical results were obtained for *Ler* and *Ler* plants carrying an active *FRI* allele (Supplementary Fig. 2).



Figure 2 Vernalization-induced changes in expression of an *FLC*–*GUS* translational fusion. A translational fusion composed of the β -glucuronidase coding sequences cloned into the *NheI* restriction site (within *FLC* exon 6) of a complementing *FLC* genomic clone was introduced into *fca-1* seedlings. *FLC*–*GUS* expression is seen throughout the non-vernalized seedling but restricted to the vasculature (strongest in cotyledon and hypocotyl but also detectable in older leaves) in vernalized seedlings.

fully establish whether they correspond exactly.

We wanted to determine whether the genotype conferring a vernalization requirement affected the histone modifications. ChIP was repeated on non-vernalized and vernalized seedlings of *Ler* and a *Ler* line transformed with an active *FRI* gene (Fig. 3). The histone modifications mirrored those observed for *fca-1* plants, so the observed changes are specific to vernalization and not to the genotype conferring the vernalization requirement, and thus occur independently of the expression level of *FLC*. We also analysed, using ChIP, the role of the *VRN* genes in determining the vernalization-induced histone modifications. The vernalization-induced changes in H3 dimethyl K9 and H3 dimethyl K4 were lost in *vrn1*—regions A and B did not show increased H3 dimethyl K9 and region E was precipitated by anti-H3 dimethyl K4. However, the

vernalization-induced H3 dimethyl K27 in region B was retained in *vrn1*. In vernalized *vrn2* seedlings, no increase in H3 dimethyl K9 in regions A and B or dimethyl K27 in region B was found; however, the vernalization-induced loss of H3 dimethyl K4 from region E was maintained or restored only weakly. Thus, both *VRN1* and *VRN2* must be required for the increase in H3 dimethyl K9 in regions A and B. *VRN1* seems to be more important than *VRN2* for the changes in region E, whereas *VRN2* is sufficient for H3 dimethyl K27 of region B. Conclusive evidence that the *VRN* proteins function directly at the *FLC* locus has yet to be obtained, but preliminary ChIP experiments using a *VRN2*–tandem affinity purification (TAP)²⁸ complementing fusion showed enrichment of *FLC* 5' sequences (region B) after immunoglobulin- γ precipitation (data not shown).

A possible model suggested by these data is that the cold causes a reduction in *FLC* transcription and this process triggers recruitment of a *VRN2* complex containing plant homologues of ESC-E(Z) to the 5' region of *FLC* (including region B), causing H3 K27 dimethylation. How cold results in low *FLC* RNA and whether any post-transcriptional regulation occurs that feeds back to cause reduced transcription is unknown at present. Nuclear run-on experiments have been undertaken, but the levels of *FLC* transcription have proved too low to be detectable. The presence of the H3 dimethyl K27 may either stimulate the methylation of H3 K9 at the 5' end of *FLC* by the ESC-E(Z) complex itself, or promote the recruitment of additional complexes that methylate H3 K9. If *VRN2* functions in an ESC-E(Z) complex similar to that in *Drosophila* and mammals, it is likely to contain one of the three E(Z) homologues, CLF, CLK and MEDEA present in the *Arabidopsis* genome²⁹. Analysis so far suggests that the vernalization response of *clf* mutant plants is not affected (R. Amasino, personal communication), so CLF is unlikely to function as the sole E(Z) component. The role of CLK and MEDEA in the *VRN2* complex and any function in vernalization remains to be established. Biochemical and genetic data show the involvement of two complexes functioning in a cooperative manner to maintain long-term gene silencing in *Drosophila* and mammals. In mammalian cells, H3 dimethyl K27 acts as a signal for the binding of POLYCOMB (PC) and recruitment of PRC1, which restricts accessibility of nucleosome remodelling factors to the chromatin⁵. PC may also stimulate the H3 K9 histone methyltransferase activity of the ESC-E(Z) complex or promote methylation of H3 K9 through the recruitment of SU(VAR)3-9 (ref. 19). A similar mechanism for long-term silencing may exist in plants; however, homologues of PC and components of the PRC1 complex are not found in the *Arabidopsis* genome sequence²⁹. The data presented here suggest that *VRN1* could mediate this long-term repression. This would predict that the function of the *VRN2* complex in methylating H3 K27 is the vernalization-specific factor shown to be required for *FLC* to be a target of *VRN1* (ref. 4). □

Methods

Plant growth conditions and vernalization treatment

Landsberg erecta and *fca-1* lines were originally obtained from M. Koornneef. *vrn1-2 fca-1* and *vrn2-1 fca-1* have been described previously^{4,15}. All seeds were surface sterilized and grown on MS media minus glucose. Vernalization was carried out by placing sterilized seeds at 4 °C under short-day conditions of 8 h white light (10 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$), 16 h dark for 6 weeks. Seeds were then transferred to long-day conditions of 16 h white light (57 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$), 8 h dark and grown for 20 days. By this stage, approximately 6 rosette leaves had been made and no flower buds were visible. Non-vernalized seeds were grown at 4 °C under short-day conditions of 8 h white light (10 $\mu\text{mol photons m}^{-2} \text{s}^{-1}$), 16 h dark for 2 days, then transferred to the long-day conditions for 20 days.

Chromatin immunoprecipitation

ChIP was carried out on 20-day-old seedlings that had or had not been vernalized using a previously described procedure²⁶. Chromatin samples were immunoprecipitated with Upstate antibodies against histone H3 dimethyl Lys 4 (catalogue no. 07-030, lot no. 22672), histone H3 dimethyl Lys 9 (catalogue no. 07-212, lot no. 23424), histone H3 dimethyl Lys 27 (catalogue no. 07-322, lot no. 22264) and histone H3 acetylation (catalogue no. 06-599, lot no. 23793). The DNA isolated from these reactions was then

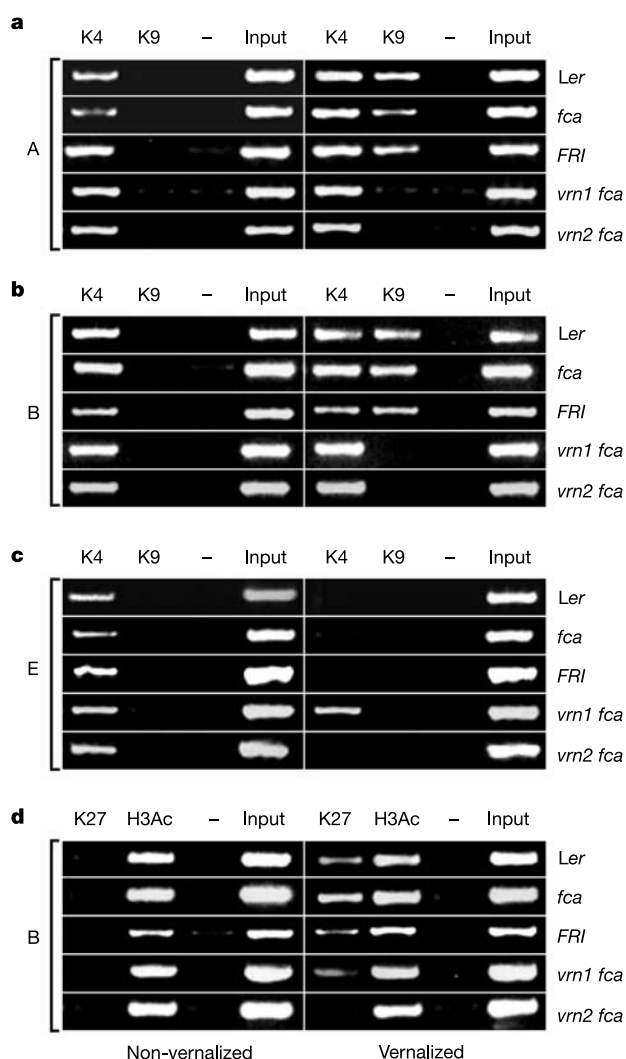


Figure 3 The histone modifications associated with vernalization are dependent on the *VRN* genes. **a, b**, Regions A and B are precipitated by the H3 dimethyl K9 antibody in vernalized wild-type seedlings, but not in vernalized *vrn1* and *vrn2* mutants. **c**, The loss of precipitation of region E with the H3 dimethyl K4 antibody is dependent on *VRN1*. Region E is precipitated from all samples of vernalized *vrn1* mutants, whereas it is only sometimes precipitated in *vrn2* mutants (in four biological replicates, low amounts of region E were precipitated in two sets; data not shown). **d**, The increased dimethylation of H3 K27 is dependent on *VRN2*. The H3 dimethyl K27 antibody precipitates region B after vernalization in wild-type and *vrn1* seedlings, but not in *vrn2*. Precipitation using the antibody to acetylated H3 is also shown.

amplified using primer pairs covering a region of *FLC* from the promoter to exon 5, a region of *FLC* previously characterized to contain *cis*-elements essential for *FLC* expression and function²⁷. Primers, shown in Supplementary Fig. 3, were designed using Primer3 software (<http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi>). Default settings were used except for product size (250–500 bp), minimum GC content (40%) and maximum GC content (60%).

Received 4 August; accepted 5 December 2003; doi:10.1038/nature02269.

- Sheldon, C. C. *et al.* The *FLF* MADS box gene: a repressor of flowering in *Arabidopsis* regulated by vernalization and methylation. *Plant Cell* **11**, 445–458 (1999).
- Michaels, S. D. & Amasino, R. M. *FLOWERING LOCUS C* encodes a novel MADS domain protein that acts as a repressor of flowering. *Plant Cell* **11**, 949–956 (1999).
- Gendall, A. R., Levy, Y. Y., Wilson, A. & Dean, C. The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. *Cell* **107**, 525–535 (2001).
- Levy, Y. Y., Mesnage, S., Mylne, J. S., Gendall, A. R. & Dean, C. Multiple roles of *Arabidopsis* *VRN1* in vernalization and flowering time control. *Science* **297**, 243–246 (2002).
- Cao, R. *et al.* Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039–1043 (2002).
- Müller, J. *et al.* Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
- Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).
- Henderson, I., Shindo, C. & Dean, C. The need for winter in the switch to flowering. *Annu. Rev. Genet.* **37**, 371–392 (2003).
- Kobayashi, Y., Kaya, H., Goto, K., Iwabuchi, M. & Araki, T. A pair of related genes with antagonistic roles in mediating flowering signals. *Science* **286**, 1960–1962 (1999).
- Kardailsky, I. *et al.* Activation tagging of the floral inducer *FT*. *Science* **286**, 1962–1965 (1999).
- Samach, A. *et al.* Distinct roles of *CONSTANS* target genes in reproductive development of *Arabidopsis*. *Science* **288**, 1613–1616 (2000).
- Lee, H. *et al.* The *AGAMOUS-LIKE 20* MADS domain protein integrates floral inductive pathways in *Arabidopsis*. *Genes Dev.* **14**, 2366–2376 (2000).
- Hepworth, S. R., Valverde, F., Ravenscroft, D., Mouradov, A. & Coupland, G. Antagonistic regulation of flowering-time gene *SOC1* by *CONSTANS* and *FLC* via separate promoter motifs. *EMBO J.* **21**, 4327–4337 (2002).
- Burn, J. E., Bagnall, D. J., Metzger, J. D., Dennis, E. S. & Peacock, W. J. DNA methylation, vernalization, and the initiation of flowering. *Proc. Natl Acad. Sci. USA* **90**, 287–291 (1993).
- Chandler, J., Wilson, A. & Dean, C. *Arabidopsis* mutants showing an altered response to vernalization. *Plant J.* **10**, 637–644 (1996).
- Birve, A. *et al.* *Su(z)12*, a novel *Drosophila* Polycomb group gene that is conserved in vertebrates and plants. *Development* **128**, 3371–3379 (2001).
- Fischle, W., Wang, Y. & Allis, C. Binary switches and modification cassettes in histone biology and beyond. *Nature* **425**, 475–479 (2003).
- Czermin, B. *et al.* *Drosophila* enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–196 (2002).
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* **16**, 2893–2905 (2002).
- Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (2001).
- Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410**, 116–120 (2001).
- Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293**, 1074–1080 (2001).
- Sims, R. J. III, Nishioka, K. & Reinberg, D. Histone lysine methylation: a signature for chromatin function. *Trends Genet.* **19**, 629–639 (2003).
- Gazzani, S., Gendall, A. R., Lister, C. & Dean, C. Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. *Plant Physiol.* **132**, 1107–1114 (2003).
- Michaels, S. D., He, Y., Scortecci, K. C. & Amasino, R. M. Attenuation of *FLOWERING LOCUS C* activity as a mechanism for the evolution of a summer-annual flowering behavior in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **100**, 10102–10107 (2003).
- Gendrel, A.-V., Lippman, Z., Jordan, C., Colot, V. & Martienssen, R. A. Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene *DDM1*. *Science* **297**, 1871–1873 (2002).
- Sheldon, C. C., Conn, A. B., Dennis, E. S. & Peacock, W. J. Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. *Plant Cell* **14**, 2527–2537 (2002).
- Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
- Goodrich, J. & Tweedie, S. Remembrance of things past: chromatin remodeling in plant development. *Annu. Rev. Cell Dev. Biol.* **18**, 707–746 (2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors thank J. Goodrich for excellent comments on the manuscript and J. Finnegan and R. Amasino for allowing us to cite their work as personal communications. This work was supported by a Core Strategic Grant to the John Innes Centre from the Biotechnology and Biological Sciences Research Council and a European Union Framework V grant to C.D.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.D. (caroline.dean@bbsrc.ac.uk).

Bcl10 activates the NF- κ B pathway through ubiquitination of NEMO

Honglin Zhou¹, Ingrid Wertz^{1,4}, Karen O'Rourke¹, Mark Utsch², Somasekar Seshagiri³, Michael Eby³, Wei Xiao⁵ & Vishva M. Dixit¹

¹Department of Molecular Oncology, ²Department of Protein Engineering, and ³Department of Molecular Biology, Genentech Inc. 1 DNA Way, South San Francisco, California 94080, USA

⁴Department of Biological Chemistry, School of Medicine, University of California at Davis, Davis, California 95616, USA

⁵Department of Microbiology and Immunology, University of Saskatchewan, Saskatoon, Saskatchewan S7N 5E5, Canada

The NF- κ B family of transcription factors is activated in response to many stimuli, including pro-inflammatory cytokines, environmental stresses and, in the case of B and T lymphocytes, by antigenic stimulation^{1,2}. Bcl10 is essential for NF- κ B activation by T- and B-cell receptors. T and B lymphocytes from Bcl10-deficient mice fail to activate NF- κ B in response to antigen-receptor stimulation and, as a consequence, are unable to proliferate³. Bcl10 overexpression is sufficient to activate NF- κ B, a process that requires the NF- κ B essential modulator NEMO (also known as IKK- γ), which is the regulatory subunit of the I κ B kinase complex⁴. However, the cellular mechanism by which Bcl10 activates the NF- κ B pathway remains unclear. Here we show that Bcl10 targets NEMO for lysine-63-linked ubiquitination. Notably, a mutant form of NEMO that cannot be ubiquitinated inhibited Bcl10-induced NF- κ B activation. Paracaspase and a ubiquitin-conjugating enzyme (UBC13) were both required for Bcl10-induced NEMO ubiquitination and subsequent NF- κ B activation. Furthermore, short interfering RNAs that reduced the expression of paracaspase and UBC13 abrogated the effects of Bcl10. Thus, the adaptor protein Bcl10 promotes activation of NF- κ B transcription factors through paracaspase- and UBC13-dependent ubiquitination of NEMO.

Activation of the 'classical' NF- κ B pathway converges on the I κ B kinase (IKK) signalosome, a protein complex composed of two kinase subunits (IKK- α and IKK- β) and a non-catalytic subunit NF- κ B essential modulator (NEMO)⁵. Activated IKK phosphorylates I κ B proteins that sequester NF- κ B transcription factors in the cytoplasm. The SCF (Skp1–Cul1–F-box) ubiquitin ligase complex then modifies the phosphorylated I κ Bs by K48-linked multiubiquitination, and thereby targets them for proteolysis in the 26S proteasome⁶. NF- κ B complexes are thus freed to translocate into the nucleus and engage κ B enhancer elements of target genes^{5,7}. Bcl10 is essential for NF- κ B activation by T- and B-cell receptors, but it is dispensable for NF- κ B activation in response to tumour-necrosis factor- α or interleukin-1 (ref. 3). Notably, Bcl10 expression is deregulated in B-cell lymphomas of mucosa-associated lymphoid tissue (MALT lymphoma) bearing the chromosomal translocation t(1; 14)(p22; q32), a rearrangement that places the *Bcl10* gene downstream of the strong immunoglobulin transcriptional enhancers. Enforced Bcl10 expression presumably contributes to transformation and antigen-independent lymphoma progression through constitutive activation of the pro-survival NF- κ B pathway^{8,9}. Bcl10 or its viral homologue E10 have been reported to interact with paracaspase/MALT1, cIAPs (inhibitor of apoptosis proteins) and TRAFs, but how these interactions affect Bcl10-induced NF- κ B activation is unclear^{10,11}.

We used a biochemical purification method to identify components of the Bcl10 signalling complex. Cytosolic extracts (S100) that were prepared from Jurkat T cells either untreated or treated with 12-O-tetradecanoylphorbol-13-acetate (TPA) plus ionomycin to mimic T-cell receptor activation were applied to a glutathione

S-transferase (GST)–Bcl10 column or, as a control, they were applied to a plain GST column. Bound proteins were then eluted and characterized by immunoblotting analysis. Endogenous Bcl10 and paracaspase bound the GST–Bcl10 column, but not the GST column, even in the absence of TPA/ionomycin stimulation (Fig. 1a); however, TPA/ionomycin treatment was required for the IKK complex to bind the GST–Bcl10 column (Fig. 1b). Binding of other proteins including TRADD, TRAF1–6, cIAP1 and cIAP2 to the GST–Bcl10 column was not detected (data not shown). Notably, the ubiquitin-conjugating E2 complex composed of UBC13 and MMS2 (also known as Uev1b) from both untreated and stimulated cells bound the GST–Bcl10 column, although the interaction was enhanced by TPA/ionomycin treatment (Fig. 1c). Because TRAF2 and TRAF6 binding to the GST–Bcl10 column was not detected,

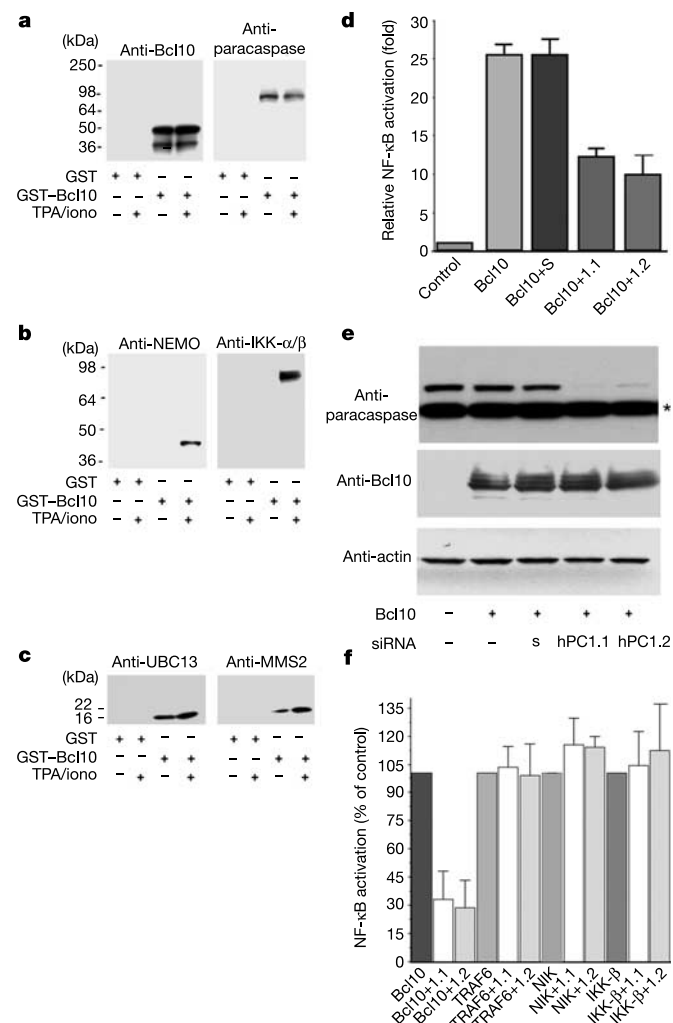


Figure 1 Isolation of a Bcl10 signalling complex and selective requirement of paracaspase for NF-κB activation by Bcl10. **a**, Western blot analysis of the elution with Bcl10 (left) or paracaspase antibody (right). **b**, Western blot analysis of the elution with NEMO (left) or IKK-α/β antibody (right). **c**, Western blot analysis of the elution with UBC13 (left) or MMS2 antibody (right). **d**, siRNA targeting paracaspase inhibits Bcl10 signalling to NF-κB activation. NF-κB activation was determined by NF-κB reporter assay. **e**, Western blot analysis of paracaspase (top) and Bcl10 (middle) expression in the indicated samples. The actin blot shows equal protein loading (bottom). The asterisk indicates a background band that serves as an internal loading control. S indicates a scrambled control siRNA. **f**, Specificity of the involvement of paracaspase in Bcl10-mediated NF-κB activation. Paracaspase siRNAs decreased paracaspase expression without affecting expression of Bcl10, TRAF6 or IKK-β in these experiments (data not shown).

UBC13–MMS2 probably associated with Bcl10 through other mechanisms^{12,13}. We confirmed the interaction of Bcl10 with paracaspase, NEMO, UBC13 and MMS2 in co-immunoprecipitation experiments on treatment with TPA/ionomycin or crosslinking antibodies to CD3 and CD28 (Supplementary Fig. 1).

We tested whether paracaspase is essential for Bcl10 to activate the NF-κB pathway using small interfering RNAs (siRNAs) that reduced paracaspase levels. 293T cells pretreated with or without siRNAs to paracaspase were transiently transfected with Bcl10, then NF-κB activation was measured by a NF-κB-responsive luciferase reporter assay. In cells transfected with either of two siRNA oligonucleotides that target paracaspase (hPC1.1, hPC1.2), Bcl10-induced NF-κB activation was significantly decreased, whereas a scrambled control siRNA had no effect (Fig. 1d). Immunoblotting confirmed that the paracaspase siRNAs caused a specific reduction in paracaspase protein levels (Fig. 1e). Paracaspase deficiency did not inhibit the effect of TRAF6- or NIK(NK-κB inducing kinase)-induced NF-κB activation, indicating that paracaspase is a specific

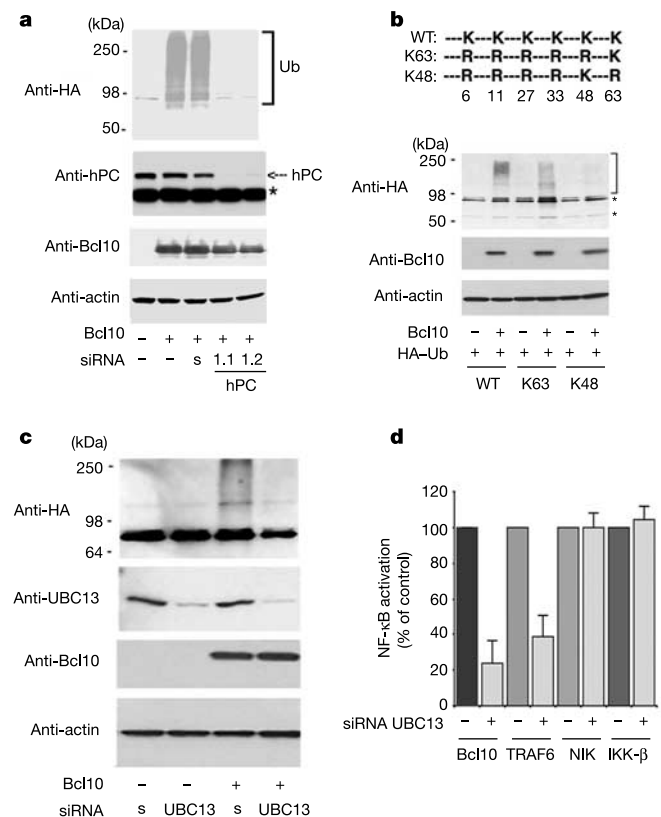


Figure 2 Paracaspase and UBC13 are required for Bcl10-induced ubiquitination and NF-κB activation. **a**, siRNA with paracaspase inhibits Bcl10-induced ubiquitination and HA-tagged ubiquitin were transfected with siRNAs targeting paracaspase or with scramble RNA. Cellular protein ubiquitination is shown by HA antibody (top). Paracaspase and Bcl10 expression was confirmed by immunoblotting. Equivalency of protein loading is shown in the actin blot (bottom). The asterisk indicates a background band that serves as an internal loading control. Ub, ubiquitin. **b**, Bcl10 promotes K63-linked ubiquitination. Bcl10 was transfected with HA-ubiquitin mutants as indicated (top panel). K63- or K48-linked ubiquitination was determined by HA immunoblotting (second panel). Bcl10 protein expression was confirmed by the anti-Bcl10 blot (third panel). The actin blot shows equal protein loading (bottom panel). WT, wild type. **c**, UBC13 is required for Bcl10-induced ubiquitination. Ubiquitination was determined by HA immunoblotting (top panel). siRNA targeting UBC13 decreased UBC13 protein expression (second panel). Bcl10 expression was confirmed (third panel). The actin blot shows equal protein loading (bottom panel). **d**, siRNA targeting UBC13 inhibits NF-κB activation by Bcl10. NF-κB activation under the indicated conditions was determined by NF-κB reporter assay.

effector of Bcl10-induced NF- κ B activation (Fig. 1f). Paracaspase siRNAs also did not inhibit NF- κ B activation by IKK- β (Fig. 1f), indicating that paracaspase functions upstream of or parallel to the IKK complex.

Because Bcl10 seemed to interact with the E2 complex UBC13-MMS2 (Fig. 1c), we explored whether Bcl10 and paracaspase are components of a ubiquitination pathway. When 293T cells were transfected with Bcl10 and haemagglutinin (HA)-tagged ubiquitin, Bcl10 increased the overall level of ubiquitination in total cell extracts (Fig. 2a). Paracaspase alone did not induce ubiquitination, although it weakly enhanced Bcl10-induced ubiquitination (data not shown). Notably, siRNA-mediated inhibition of endogenous paracaspase expression abrogated Bcl10-induced cellular ubiquiti-

nation (Fig. 2a). Because paracaspase seems to be required for both NF- κ B activation and ubiquitination induced by Bcl10, we investigated whether the two phenomena might be linked.

The fate of a protein that has been modified by ubiquitination is determined by the type of ubiquitin linkage. K48-linked polyubiquitination generally targets proteins for degradation in the proteasome, whereas K63-linked ubiquitination can regulate protein function or protein-protein interactions¹⁴. When Bcl10 was co-expressed with HA-tagged ubiquitin or ubiquitin mutants with only K48 or K63 available for polymerization, it induced K63-linked but not K48-linked ubiquitination (Fig. 2b). As UBC13 and MMS2

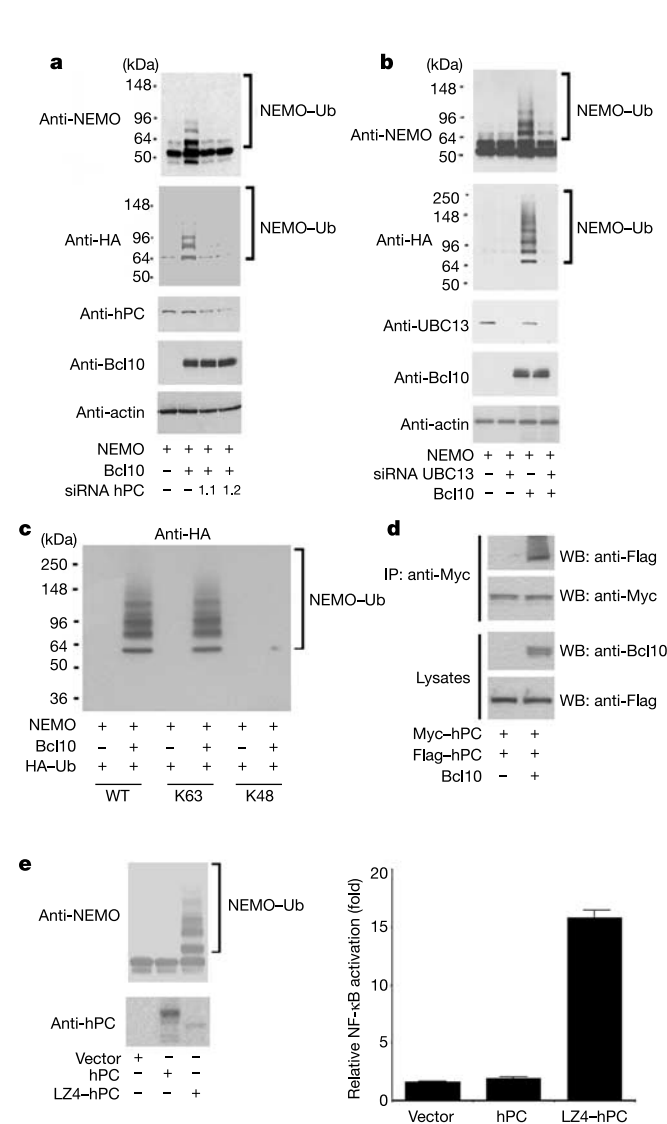


Figure 3 NEMO is a substrate of the Bcl10-induced ubiquitination pathway. **a**, Paracaspase is required for Bcl10-induced NEMO ubiquitination. NEMO ubiquitination is detected by anti-NEMO and anti-HA blots. The actin blot shows equal protein loading. **b**, UBC13 is required for Bcl10-induced NEMO ubiquitination. NEMO ubiquitination is detected by anti-NEMO and anti-HA blots. The actin blot shows equal protein loading (bottom). **c**, Bcl10 induces K63-linked NEMO ubiquitination. K63- or K48-linked ubiquitination was determined by anti-HA blot (right). NEMO expression was confirmed (left). Bcl10 expression was also confirmed (data not shown). **d**, Bcl10 induces oligomerization of paracaspase. Myc-hPC and Flag-hPC were transfected into 293T cells with or without Bcl10. Immunoprecipitation with anti-Myc detected Flag-hPC in the presence but not absence of Bcl10. **e**, Enforced oligomerization of paracaspase is sufficient for NEMO ubiquitination and NF- κ B activation.

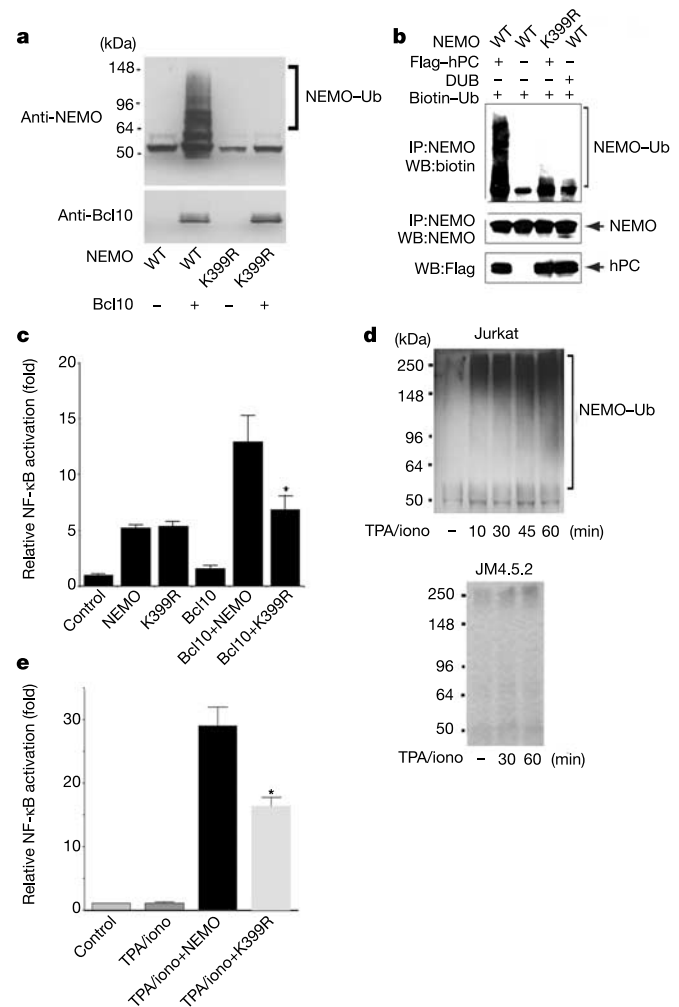


Figure 4 Role of NEMO ubiquitination in NF- κ B activation. **a**, Ubiquitination modification site of NEMO. NEMO or the NEMO mutant (K399R) together with the indicated plasmids were transfected into 293T cells. NEMO ubiquitination was determined by anti-NEMO immunoblotting. **b**, Paracaspase-mediated *in vitro* ubiquitination of wild-type but not K399R NEMO. NEMO ubiquitination was determined by NEMO immunoprecipitation (IP) and immunoblotting of biotinylated ubiquitin. DUB indicates deubiquitinating treatment with Usp2. **c**, Complementations of NEMO-deficient 5R cells with NEMO or NEMO K399R mutant. 5R cells were co-transfected with Bcl10 and NEMO or the NEMO mutant. Bcl10-stimulated NF- κ B activation was determined by NF- κ B reporter assay. Asterisk, $P < 0.02$; Student's *t*-test. **d**, TPA/ionomycin treatment stimulated NEMO ubiquitination. Extracts from TPA/ionomycin-treated Jurkat cells or JM4.5.2 cells were immunoprecipitated with NEMO antibody followed by immunoblotting with ubiquitin antibody. **e**, Complementations of the NEMO-deficient JM4.5.2 cell with wild-type NEMO or NEMO K399R mutant. JM4.5.2 cells were transfected with NEMO or the NEMO K399R mutant. TPA/ionomycin-stimulated NF- κ B activation was determined by NF- κ B reporter assay. Asterisk, $P < 0.05$; Student's *t*-test.

form an E2 complex that functions with certain E3 ubiquitin ligases such as TRAF2, TRAF6, or RAD5 to catalyse K63-linked polyubiquitination^{12–16}, we tested whether UBC13 is required for Bcl10-induced ubiquitination and NF- κ B activation. 293T cells were co-transfected with Bcl10 and siRNAs targeting UBC13. Knockdown of UBC13 protein was found to abrogate Bcl10-induced ubiquitination (Fig. 2c). Furthermore, consistent with the hypothesis that Bcl10-induced ubiquitination is linked to NF- κ B activation, UBC13 siRNA also inhibited Bcl10-mediated NF- κ B activation (Fig. 2d). UBC13 siRNAs also inhibited ubiquitination and NF- κ B activation induced by TRAF6 (Fig. 2d, data not shown), consistent with previous findings using a dominant-negative UBC13 mutant¹³. Knockdown of UBC13 did not inhibit NF- κ B activation by IKK- β or NIK (Fig. 2d). Immunoblot experiments revealed that siRNAs targeting UBC13 reduced its expression, but not expression of Bcl10, TRAF6, NIK or IKK- β (data not shown).

To explore the link between Bcl10-induced ubiquitination and NF- κ B activation, we examined the ubiquitination status of proteins that have an impact on NF- κ B signalling. Only NEMO became increasingly ubiquitinated when coexpressed with Bcl10—CARD11, TRAF1–6, cIAP1, cIAP2, IKK- α and IKK- β were not ubiquitinated in response to Bcl10 overexpression (Fig. 3a; data not shown). This Bcl10-induced ubiquitination of NEMO involved K63 rather than K48 linkages and was decreased by siRNAs targeting either paracaspase or UBC13 (Fig. 3a–c). One potential mechanism by which Bcl10 induces paracaspase- and UBC13-dependent ubiquitination of NEMO is through binding to paracaspase and inducing conformational changes that lead to the assembly of a ubiquitin ligase complex. Consistent with a role for Bcl10 in mediating paracaspase oligomerization, Myc-tagged paracaspase was co-immunoprecipitated with Flag-tagged paracaspase only in the presence of coexpressed Bcl10 (Fig. 3d). In addition, fusion of the caspase-like domain of paracaspase to a tetramer-forming leucine zipper domain (LZ4–hPC) resulted in NEMO ubiquitination and NF- κ B activation in the absence of coexpressed Bcl10 (Fig. 3e).

To determine whether paracaspase was associated with ubiquitin ligase activity, we purified Flag-tagged versions of paracaspase from baculovirus-infected Hi5 insect cells and performed *in vitro* ubiquitination assays. Full-length paracaspase and the carboxy terminus alone, but not the amino terminus, exhibited *in vitro* autoubiquitination (Supplementary Fig. 2a). These results indicate that a domain within the paracaspase C terminus or associated proteins possesses ubiquitin ligase activity. A maltose-binding protein (MBP)-tagged version of the paracaspase C terminus was purified from bacteria, which lack the ubiquitin system, and this paracaspase fusion was still capable of *in vitro* autoubiquitination (Supplementary Fig. 2b). This result suggests that the caspase-like domain of paracaspase, rather than an associated insect protein, had intrinsic ubiquitin ligase activity.

Deletion mutants of NEMO were used to identify which lysine residue was modified with ubiquitin. Bcl10 failed to induce ubiquitination of a NEMO C-terminal deletion mutant missing amino acids 394–412 (data not shown). This region of NEMO contains a putative zinc finger motif which is required for full IKK activation^{17,18}. As ubiquitination of this NEMO mutant may be prevented by disrupted interactions with other molecules, we tested full-length NEMO in which the single lysine residue between amino acids 394–412 was mutated to an arginine (K399R). Bcl10 did not promote ubiquitination of the NEMO K399R mutant, whereas immunoblotting confirmed that the NEMO K399R mutant was expressed at a similar level to wild-type NEMO (Fig. 4a). These results suggest that K399 is the acceptor lysine on NEMO for Bcl10-induced ubiquitination. Indeed, paracaspase-mediated *in vitro* ubiquitination of NEMO K399R was greatly attenuated relative to wild-type NEMO (Fig. 4b). Consistent with a critical role for ubiquitination in IKK activation, the K399 modification site of

human NEMO is conserved in NEMO from many other species (Supplementary Fig. 3).

To determine whether NEMO ubiquitination is necessary for Bcl10-induced NF- κ B activation, either wild-type NEMO or NEMO K399R was co-transfected with Bcl10 into NEMO-deficient 5R cells¹⁹. Whereas Bcl10 alone failed to activate NF- κ B in the 5R cells, NF- κ B activation did occur in 5R cells transfected with wild-type NEMO. NEMO K399R permitted much weaker NF- κ B activation (Fig. 4c). These results indicate that NEMO ubiquitination is important for full NF- κ B activation by Bcl10.

Because Bcl10 normally acts downstream of B- and T-cell receptors, we determined whether endogenous NEMO is ubiquitinated in Jurkat T cells. Using TPA/ionomycin to mimic T-cell-receptor signalling, NEMO ubiquitination was detected within 10 min of treatment (Fig. 4d). No degradation of ubiquitinated NEMO was detected over time (data not shown), consistent with ubiquitination altering NEMO function rather than targeting it for degradation. To test whether NEMO ubiquitination is necessary for NF- κ B activation in Jurkat T cells, wild-type NEMO or NEMO K399R was expressed in the NEMO-deficient Jurkat cell line JM5.4.2 (ref. 20). TPA/ionomycin failed to activate a NF- κ B-dependent reporter gene in the NEMO-deficient parental cells, but NF- κ B activity was observed when the cells were transfected with wild-type NEMO. Expression of NEMO K399R in the JM5.4.2 cells permitted weaker NF- κ B activation after TPA/ionomycin stimulation (Fig. 4e). These results indicate that NEMO ubiquitination has an important role in antigen-receptor-stimulated NF- κ B activation.

We have shown that Bcl10-mediated activation of the NF- κ B pathway requires a paracaspase-dependent ubiquitination event. Paracaspase lacks a RING or HECT domain normally found in ubiquitin ligases^{14,21}, but instead uses its C-terminal caspase-like domain to catalyse ubiquitination (Supplementary Fig. 2a, b). Bcl10 probably activates paracaspase by inducing conformational changes that permit its oligomerization and formation of a complex with UBC13–MMS2. We speculate that K63-linked multiubiquitination of NEMO either recruits an upstream regulator such as an IKK kinase to activate the IKKs, or induces a conformational change within the IKK complex that leads to IKK autophosphorylation and activation²². Recent studies show that the tumour suppressor gene product CYLD negatively regulates NF- κ B activation by acting as a deubiquitinating enzyme^{23–25}, so it is tempting to speculate that ubiquitination of NEMO represents a prevalent regulatory mechanism by which divergent stimuli activate the NF- κ B pathway. □

Methods

Cell lines, reagents and antibodies

HEK 293T cells and Jurkat cells were cultured as previously described¹¹. The JM.4.5.2 cell line was a gift from S.-C. Sun and the 5R cell line was a gift from S. Yamaoka. Usp2 was a gift from R. Baker. UBC13 and MMS2 antibodies were raised against purified UBC13 and human MMS2 proteins in BALB/c mice. MMS2 antibody reacts with both MMS2 and Uev1 equally well in western blot analysis. Paracaspase monoclonal antibodies have been described previously¹¹. All other antibodies were from Santa Cruz Biotechnology. The construct LZ4–hPC expresses the paracaspase caspase-like domain (298–814) as a fusion with a tetramer-forming leucine zipper domain²⁶. NF- κ B luciferase reporter assays were measured with the Dual Luciferase Assay System (Promega). The following reagents were purchased as indicated: Fugene 6 (Roche), TPA and ionomycin (Sigma), Lipofectamine 2000 (Invitrogen), Site-directed mutagenesis kit (Stratagene), siRNA oligonucleotides (Dharmacon).

Identification of Bcl10-binding proteins

Human Jurkat cells were treated for 30 min with or without 50 ng TPA plus 50 ng ionomycin. S100 cell extracts were prepared as previously described and pre-washed with the GST column to eliminate nonspecific protein binding before loading on the GST–Bcl10 affinity column. Binding and elution of proteins as well as immunoprecipitation with Bcl10 antibody were performed as previously described^{11,27}.

Transient transfection and NF- κ B reporter assays

HEK 293T cells were transfected as previously described¹¹. Twenty-four hours after transfection, NF- κ B luciferase activity was measured using the Dual Luciferase Reporter Assay system (normalized to control luciferase activity). JM.4.5.2 cells were transfected

with Fugene 6. At 48 h after transfection, cells were treated with or without 50 ng ml⁻¹ TPA plus 1 μM ionomycin. NF-κB reporter activity was measured after 6 h treatment.

Protein purification

Human Bcl10, UBC13, MMS2 and NEMO were cloned into GST fusion vector pGEX6P (Amersham Pharmacia Biotech) and sequence verified. Proteins expressed in the *Escherichia coli* strain BL21(DE3)-RIL (Stratagene) were purified either as previously described²⁸ or according to the manufacturer's instructions. Cleavage with PreScission Protease (Amersham Pharmacia Biotech) results in the separation of GST from UBC13, MMS2, or NEMO. Paracaspase (amino acids 312–582) was cloned into the pMal-p2G vector (New England Biolabs). The MBP–paracaspase fusion protein was expressed and purified according to the manufacturer's instructions. We purified N-terminal Flag-tagged recombinant proteins using anti-Flag-M2 affinity gel (Sigma) from Hi5 cells infected with the recombinant viruses expressing full-length (residues 2–824), N-del (residues 348–824) or C-del (residues 2–347) paracaspase.

RNA interference

293T cells in 6 cm dishes were transfected with 21-nucleotide complementary RNAs with symmetrical two-nucleotide overhangs using Lipofectamine 2000. For siRNA targeted to paracaspase, hPC1.1 (AACCCAGAGTCAAAGGCAGTC) and hPC1.2 (AAGGCAGTCTTGGCTGGACAG) were used. For siRNA targeted to UBC13, four oligonucleotides (TCAAAGCCGAACAGATGA, AAGTACGTTTCATGACCAA, TCCTAATGTAGACAAGTTG and GATCCGCACAGTCTTGCTA) were mixed for usage. At 48 h after adding siRNAs, cells were transfected with expression plasmids and cultured for an additional 24 h before experiments.

In vitro ubiquitination assay

In vitro ubiquitination assays were performed in a 50-μl reaction volume containing the following: 2 μg N-terminal biotinylated ubiquitin (Boston Biochem), 0.2 μg E1 (Calbiochem), 0.5 μg each of UBC13 and MMS2, 1 μg Flag-purified paracaspase or 1 μg MBP–paracaspase (312–582) fusion protein with 1 M Na citrate, 5 μl × 10 reaction buffer (300 mM HEPES pH 7.2, 20 mM ATP, 50 mM MgCl₂, 2 mM dithiothreitol), and 100 ng recombinant NEMO. Reactions were incubated at 30 °C for 1 h. Autoubiquitination reaction was terminated by addition of sample buffer or alternatively, NEMO ubiquitination reactions were diluted to 1 ml with buffer containing 50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA and 1 mM dithiothreitol, and NEMO immunoprecipitated from the reactions. Where indicated, the immunoprecipitates were further incubated with 2.5 μg Usp2 for 1 h at 25 °C. Proteins were separated by SDS–polyacrylamide gel electrophoresis, then immunoblotted with the indicated antibodies.

Immunoprecipitation

Immunoprecipitation with NEMO antibody was performed after dissociation of samples. Cells were lysed in immunoprecipitation buffer (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, protease inhibitor cocktail) plus 1% SDS and boiled for 5 min at 95 °C. Supernatants were diluted tenfold with immunoprecipitation buffer, and NEMO antibody was added to perform the immunoprecipitation as described previously.

Received 20 October; accepted 2 December 2003; doi:10.1038/nature02273.

Published online 21 December 2003.

- Silverman, N. & Maniatis, T. NF-κB signaling pathways in mammalian and insect innate immunity. *Genes Dev.* **15**, 2321–2342 (2001).
- Dixit, V. M. & Mak, T. W. NF-κB signaling: Many roads lead to Madrid. *Cell* **111**, 615–619 (2002).
- Ruland, J. *et al.* Bcl10 is a positive regulator of antigen receptor-induced activation of NF-κB and neural tube closure. *Cell* **104**, 33–42 (2001).
- Lucas, P. C. *et al.* Bcl10 and MALT1, independent targets of chromosomal translocation in malt lymphoma, cooperate in a novel NF-κB signaling pathway. *J. Biol. Chem.* **276**, 19012–19019 (2001).
- Rothwarf, D. M. & Karin, M. The NF-κB activation pathway: a paradigm in information transfer from membrane to nucleus. *Sci. STKE* **5**, RE1 (1999).
- Ben-Neriah, Y. Regulatory functions of ubiquitination in the immune system. *Nature Immunol.* **3**, 20–26 (2002).
- Hoffmann, A., Levchenko, A., Scott, M. L. & Baltimore, D. The IκB–NF-κB signaling module: temporal control and selective gene activation. *Science* **298**, 1241–1245 (2002).
- Willis, T. G. *et al.* Bcl10 is involved in t(1;14)(p22;q32) of MALT B cell lymphoma and mutated in multiple tumor types. *Cell* **96**, 35–45 (1999).
- Zhang, Q. *et al.* Inactivating mutations and overexpression of BCL10, a caspase recruitment domain-containing gene, in MALT lymphoma with t(1;14)(p22;q32). *Nature Genet.* **22**, 63–68 (1999).
- Thome, M. & Tschopp, J. TCR-induced NF-κB activation: a crucial role for Carma1. *Bcl10 and MALT1. Trends Immunol.* **24**, 419–424 (2003).
- Uren, A. G. *et al.* Identification of paracaspases and metacaspases: two ancient families of caspase-like proteins, one of which plays a key role in MALT lymphoma. *Mol. Cell* **6**, 961–967 (2000).
- Shi, C. S. & Kehrli, J. H. Tumor necrosis factor (TNF)-induced germinal center kinase-related (GCKR) and stress-activated protein kinase (SAPK) activation depends upon the E2/E3 complex Ubc13-Uev1A/TNF receptor-associated factor 2 (TRAF2). *J. Biol. Chem.* **278**, 15429–15434 (2003).
- Deng, L. *et al.* Activation of the IκB kinase complex by TRAF6 requires a dimeric ubiquitin-conjugating enzyme complex and a unique polyubiquitin chain. *Cell* **103**, 351–361 (2000).
- Weissman, A. M. Themes and variations on ubiquitylation. *Nature Rev. Mol. Cell Biol.* **2**, 169–178 (2001).
- Wang, C. *et al.* TAK1 is a ubiquitin-dependent kinase of MKK and IKK. *Nature* **412**, 346–351 (2001).
- Hoegge, C. *et al.* RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419**, 135–141 (2002).
- Makris, C., Roberts, J. L. & Karin, M. The carboxyl-terminal region of IκB kinase gamma (IKKγ) is required for full IKK activation. *Mol. Cell. Biol.* **22**, 6573–6581 (2002).

- Smahi, A. *et al.* Genomic rearrangement in NEMO impairs NF-κB activation and is a cause of incontinentia pigmenti. *Nature* **405**, 466–472 (2000).
- Yamaoka, S. *et al.* Complementation cloning of NEMO, a component of the IκB kinase complex essential for NF-κB activation. *Cell* **93**, 1231–1240 (1998).
- Harhaj, E. W. *et al.* Somatic mutagenesis studies of NF-κB signaling in human T cells: evidence for an essential role of IKKγ in NF-κB activation by T-cell costimulatory signals and HTLV-I Tax protein. *Oncogene* **19**, 1448–1456 (2000).
- Ohi, M. D. *et al.* Structural insights into the U-box, a domain associated with multi-ubiquitination. *Nature Struct. Biol.* **10**, 250–255 (2003).
- Ghosh, S. & Karin, M. Missing pieces in the NF-κB puzzle. *Cell* **109**, S81–S96 (2002).
- Kovalenko, A. *et al.* The tumour suppressor CYLD negatively regulates NF-κB signalling by deubiquitination. *Nature* **424**, 801–805 (2003).
- Brummelkamp, T. *et al.* Loss of the cylindromatosis tumour suppressor inhibits apoptosis by activating NF-κB. *Nature* **424**, 797–801 (2003).
- Trompouki, E. *et al.* CYLD is a deubiquitinating enzyme that negatively regulates NF-κB activation by TNFR family members. *Nature* **424**, 793–796 (2003).
- Harbury, P. B. *et al.* A switch between two-, three- and four-stranded coiled coil in GCN4 leucine zipper mutants. *Science* **262**, 1401–1407 (1993).
- Li, M. *et al.* Deubiquitination of p53 by HAUSP is an important pathway for p53 stabilization. *Nature* **416**, 648–653 (2002).
- Moraes, T. F. *et al.* Crystal structure of the human ubiquitin conjugating enzyme complex, hMms2–hUbc13. *Nature Struct. Biol.* **8**, 669–673 (2001).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank members of the Dixit laboratory for discussions and support; K. Newton for critical reading of the manuscript; S. Sun for the JM4.5.2 cell line; S. Yamaoka for the 5R cell line; R. Baker for Usp2; M. Yan for experimental advice; Z. Zhang for bioinformatics analysis; P. Andersen for antibody preparation; P. Yin for protein purification; and the Genentech Sequencing Core. I.E.W. is funded by a PSTP fellowship from the University of California at Davis.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.M.D. (dixit@gene.com).

Vinculin activation by talin through helical bundle conversion

Tina Izard¹, Gwyndaf Evans³, Robert A. Borgon^{1,4}, Christina L. Rush^{1,4}, Gerard Bricogne³ & Philippe R. J. Bois²

¹Department of Hematology-Oncology, and ²Department of Genetics, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA

³Global Phasing Limited, Sheraton House, Castle Park, Cambridge CB3 0AX, UK

⁴Department of Molecular Sciences, University of Tennessee, Memphis, Tennessee 38163, USA

Vinculin is a conserved component and an essential regulator of both cell–cell (cadherin-mediated) and cell–matrix (integrin–talin-mediated focal adhesions) junctions, and it anchors these adhesion complexes to the actin cytoskeleton by binding to talin in integrin complexes or to α-actinin in cadherin junctions^{1–3}. In its resting state, vinculin is held in a closed conformation through interactions between its head (Vh) and tail (Vt) domains^{4–6}. The binding of vinculin to focal adhesions requires its association with talin. Here we report the crystal structures of human vinculin in its inactive and talin-activated states. Talin binding induces marked conformational changes in Vh, creating a novel helical bundle structure, and this alteration actively displaces Vt from Vh. These results, as well as the ability of α-actinin to also bind to Vh and displace Vt from pre-existing Vh–Vt complexes, support a model whereby Vh functions as a domain that undergoes marked structural changes that allow vinculin to direct cytoskeletal assembly in focal adhesions and adherens junctions. Notably, talin's effects on Vh structure establish helical bundle conversion as a signalling mechanism by which proteins direct cellular responses.

Vinculin's head and tail domains are the most conserved portions

of the protein, and their interaction masks cryptic binding sites for talin and α -actinin to Vh, and for F-actin and paxillin to Vt^{5,7–9}. Thus, vinculin activation has long been speculated to require structural changes that sever the Vh–Vt interaction. Key mediators of this process have been suggested to be acidic phospholipids such as phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)P₂), which can bind to and alter the conformation of Vt^{5,10} and facilitate talin binding to Vh⁵. In focal adhesions talin binds to Vh through three high-affinity vinculin-binding sites (VBS1, VBS2 and VBS3) present in its central rod domain¹¹. Until quite recently talin was thought to function simply as a bridge between vinculin and integrin receptors, yet talin has also been suggested to trigger signalling and vinculin activation in focal adhesions^{11–13}. To define the mechanism of vinculin activation in adhesion complexes and the role of talin in this process, we determined the crystal structures of the human Vh–Vt complex (residues 1–258 and 897–1066) at 2.35 Å resolution, and of the Vh domain in complex with talin VBS3 at 2.7 Å.

The crystal structure of chicken Vt revealed five amphipathic α -helices (H1–5) that form an anti-parallel bundle, and this was suggested to represent its inactive form, as its conformation changed on binding to PtdIns(4,5)P₂¹⁰. Our human Vt structure in complex with Vh indeed confirms the closed conformation of Vt. Figure 1 shows a superposition of two chicken Vt molecules in the asymmetric unit onto our human Vt as seen in complex with Vh. Excluding the termini and the flexible loop following H4 (residues 895–1003 and 1014–1045), the C α positions of human Vh can be superimposed onto the 144 equivalent C α positions of chicken Vt with root-mean-square deviation (r.m.s.d.) of 0.88 and 0.91 Å for either molecule in the asymmetric unit. The only significant

differences between human and chicken Vt are observed in the loop regions, where the short amino-terminal helix H0 present in chicken Vt is a random coil in our human Vt complexed structure, and the loops connecting helices H4 and H5 and that following helix H5 are disordered or display a different conformation. The recent structure determination of the cytoskeletal protein α -catenin^{14–16} reveals that its M fragment has the highest structural similarity (<http://www.ebi.ac.uk/dali>) to our human Vt structure, as discussed previously¹⁵.

The structure of the Vh domain was heretofore unknown. The crystal structure revealed that in its inactive conformation Vh adopts an elongated conformation (101 × 22 × 46 Å, Fig. 2). Vh comprises seven amphipathic α -helices (α 1–7) arranged as two four-helical bundles; that is, the N-terminal bundle (α 1–4) and the carboxy-terminal bundle (α 4–7). The centrally positioned long α 4 helix (residues 102–147) bridges the two substructures. The constellation of the helices in each bundle is in an anti-parallel up-down-up-down configuration. The last two amino acids of helix α 2 are followed by five residues that adopt a helical-like structure, but their temperature factors, and those on the following loop, are higher (up to 87 Å² for the main chain) than those in the remainder of the molecule. The C-terminal half of α 4 is nearly parallel with α 7, whereas its N-terminal half is nearly parallel with α 3. In the C-terminal bundle helices α 6 and α 5 are nearly parallel and an invariant proline residue in both induces a kink. The overall topology places the C-terminal bundle at a 20° angle relative to the N-terminal bundle.

Several structures with helical bundle architecture are known, yet the closed conformation of Vh shows remarkable similarity to the dimerization domain of α -catenin^{14,15}. α -Catenin links the cadherin– β -catenin complex in adherens junctions to the cytoskeleton through its interactions with α -actinin and vinculin^{17,18}. In its resting state α -catenin is a homodimer, with each protomer containing five amphipathic α -helices that form two distinct substructures¹⁴. Notably, the binding of an amphipathic α -helix from β -catenin displaces the homodimer to form a seven- α -helix

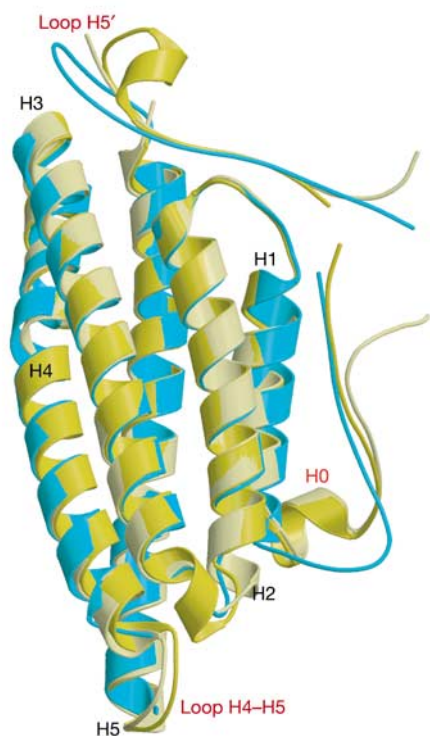


Figure 1 The closed conformation of the human vinculin tail. Superposition of inactive human Vt (residues 879–1066; light blue) in its Vh-bound conformation onto either chicken Vt monomer in the asymmetric unit¹⁰ (yellow and light yellow). Labelling of secondary structure elements, helices H1–5, corresponds to those of chicken Vt (PDB 1qkr). The largest discrepancies between the two structures are found in the loop regions, which are labelled in red. Loop H5' denotes the loop following helix H5.

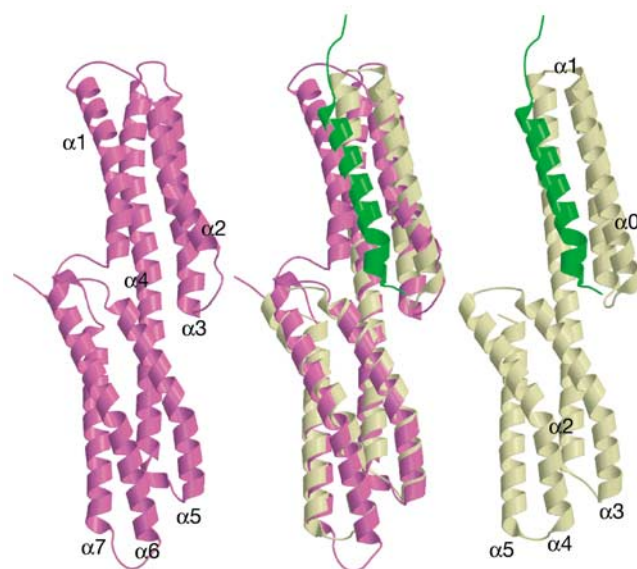


Figure 2 Crystal structure of inactive human Vh (pink, residues 1–258; left panel) compared to α -catenin (light yellow, residues 57–261) bound to β -catenin (green, residues 118–149 (PDB 1dow); right panel). The middle panel shows the superposition of Vh onto the α -catenin– β -catenin¹⁴ complex. The seven α -helices present in Vh are labelled α 1–7. Helix α 1 of Vh is structurally equivalent to the β -catenin helix; Vh helices α 2–7 correspond to α 0– α 5 of α -catenin. In Vh, helices α 4 and α 7 are followed by a short 3₁₀ helix.

structure¹⁴ that is markedly similar to the structure of Vh. Specifically, the β -catenin and α 0 α -catenin helices are structurally highly related to Vh α 1 and α 2, whereas α -catenin helices α 1–5 are nearly equivalent to those of Vh α 3–7 (Fig. 2). The C α positions comprising Vh helices α 3–7 can be superimposed onto the equivalent C α positions of α -catenin helices α 1–5 of the β -catenin: α -catenin structure with a r.m.s.d. of 1.44 Å. The helical exchange observed after the binding of β -catenin to α -catenin thus suggested that a similar mechanism might apply during the binding of Vh to its cytoskeletal-binding partners.

Vh has a high affinity for Vt (dissociation constant, K_d , of 50–80 nM)^{19,20}. Indeed, the crystal structure of the Vh–Vt complex (Fig. 3a) reveals that eight hydrogen bonds (from seven invariant residues) are formed across the interface. Invariant Glu 28 is sandwiched between invariant Lys 944 and conserved Arg 945, and 26 hydrophobic residues (18 invariant) account for van der Waals interdomain interactions (Fig. 3b). The residues responsible for this extensive interface reside on Vh helices α 1 and α 4, and on Vt helices H3, H4 and H5 (Supplementary Table S1). A strong shape complementarity across this interface was also confirmed by calculating the shape correlation statistic²¹, which was 0.762, a value corresponding to those observed for protein oligomeric interfaces. Moreover, almost 9% (approximately 1,300 Å) and 14% (approximately 1,400 Å) of the entire solvent-accessible surface areas of Vh and Vt respectively are buried in the complex, values that are higher than those occurring in the α -catenin– β -catenin complex¹⁴. A region buried by the Vh–Vt interaction (residues 978–1000) binds to paxillin (Supplementary Fig. S1)⁹. Similarly, several residues within a region (1001–1028) that targets other focal adhesion proteins^{6,8} are buried on Vh–Vt complex formation, again suggesting a competition for binding sites. Finally, the Vh–Vt complex creates a large acidic pocket that is lined by several aspartate, glutamate and glutamine residues residing on α 4 and the C terminus of Vh, as well as on H5 of Vt (Fig. 3c).

The binding of talin to vinculin has been suggested to inhibit the Vh–Vt interaction and to expose actin-binding sites in Vt^{20,22}. Talin VBS3 (residues 1944–1969) binds Vh with high affinity (K_d of 39 nM)²⁰. Indeed in our Vh–VBS3 crystal structure one molecule of

VBS3 binds to one molecule of Vh in an intimate fashion, as VBS3 is almost swallowed by Vh, burying 46% of its solvent-accessible surface area (Fig. 4).

Unexpectedly, the Vh–VBS3 crystal structure reveals striking changes in the structure of the N-terminal bundle of Vh on binding to talin. As a net result, all four of these α -helices of Vh completely rearrange to make room for talin VBS3. Notably, the resulting structure of the new five-helical bundle establishes helical conversion in Vh, whereby Vh helices α 1–4 are transformed into a totally unique α -helical arrangement (Fig. 4b, c; see also Supplementary Fig. S2).

Talin VBS3 forms one amphipathic α -helix that inserts between helices α 1 and α 2 of Vh. Binding occurs largely through van der Waals interactions, whereby the hydrophobic face of VBS3 interacts with the hydrophobic core of the N-terminal helical bundle of Vh. Eighteen Vh residues (seven on α 2, six on α 1 and four on α 4) are in van der Waals contact with 13 side chains of VBS3 (Supplementary Table S2). Helix α 3 is positioned furthest away from the talin-binding (and the Vt-binding) site, and thus generally maintains its structure, moving only a few Ångströms when Vh is bound by VBS3. However, Vh helices α 1, α 2 and α 4 take on new identities in the Vh–VBS3 complex. The loop connecting Vh helices α 1 and α 2 moves over 15 Å to wrap around the C-terminal end of talin VBS3. This displacement is accomplished by a rotation about the C terminus of α 2 of over 24° as well as by a sharp bend at Pro 15 within α 1 to kink this helix by 40°. Next, talin VBS3 distorts the N-terminal half of α 4, which pivots almost 20° from its inactive position. Finally, talin binding also provokes a prominent kink at Gly 118 in α 4 and this shifts the N-terminal residue of α 4 by over 12 Å.

Despite these marked alterations at its N terminus, the C-terminal half of the shared α 4 helix that extends into the C-terminal bundle of Vh retains the structure of the inactive conformation. The C-terminal bundle of Vh is not in contact with talin VBS3, nor is this bundle altered in its structure (the r.m.s.d. of the C α positions of VBS3-bound versus Vt-bound Vh residues 150–250 is 0.75 Å). Thus, the C-terminal helical bundle of Vh functions as an anchor, providing a rigid scaffold for the movements occurring during

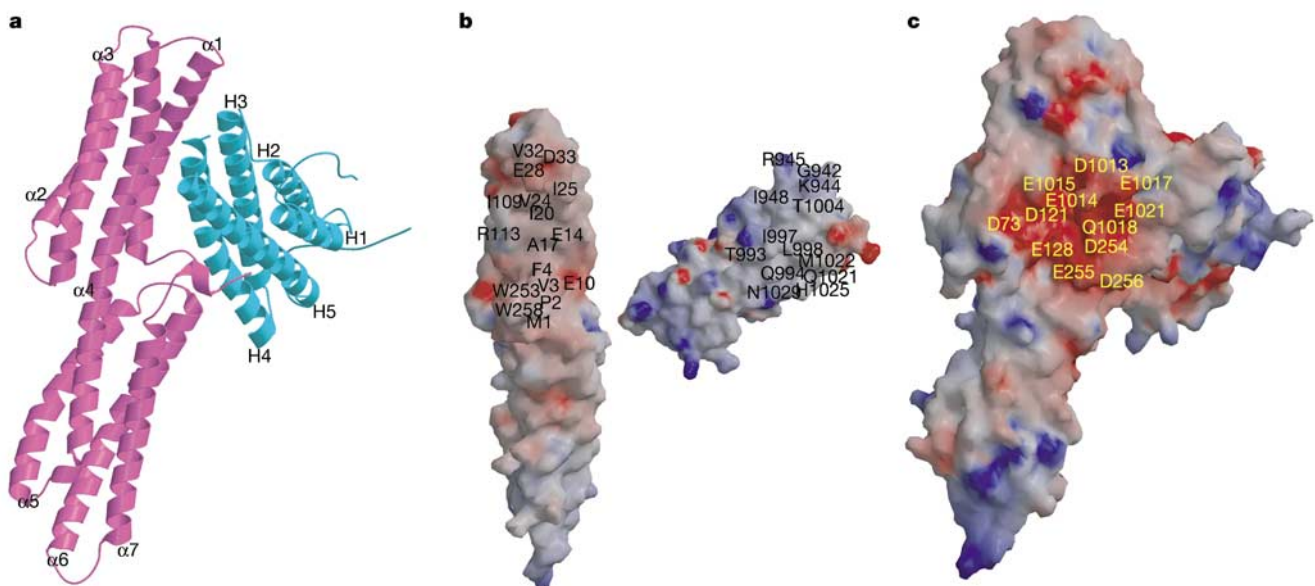


Figure 3 Structure of inactive human vinculin. **a**, Cartoon representation of the closed conformation of vinculin (Vh, residues 1–258, pink; Vt, 879–1066, light blue). **b**, **c**, Electrostatic surface potential (red, negative; blue, positive) of the Vh–Vt complex. **b**, Head-on view of each interface (left, Vh; right, Vt) when Vh and Vt are taken apart (that

is, Vh is rotated 90° to the left and Vt is rotated 90° to the right with respect to the orientation shown in **a**). Residues involved in interdomain contacts are labelled. **c**, Same orientation as in **a**, revealing the acidic pocket created when Vh binds to Vt. Acidic residues lining the pocket are indicated in yellow.

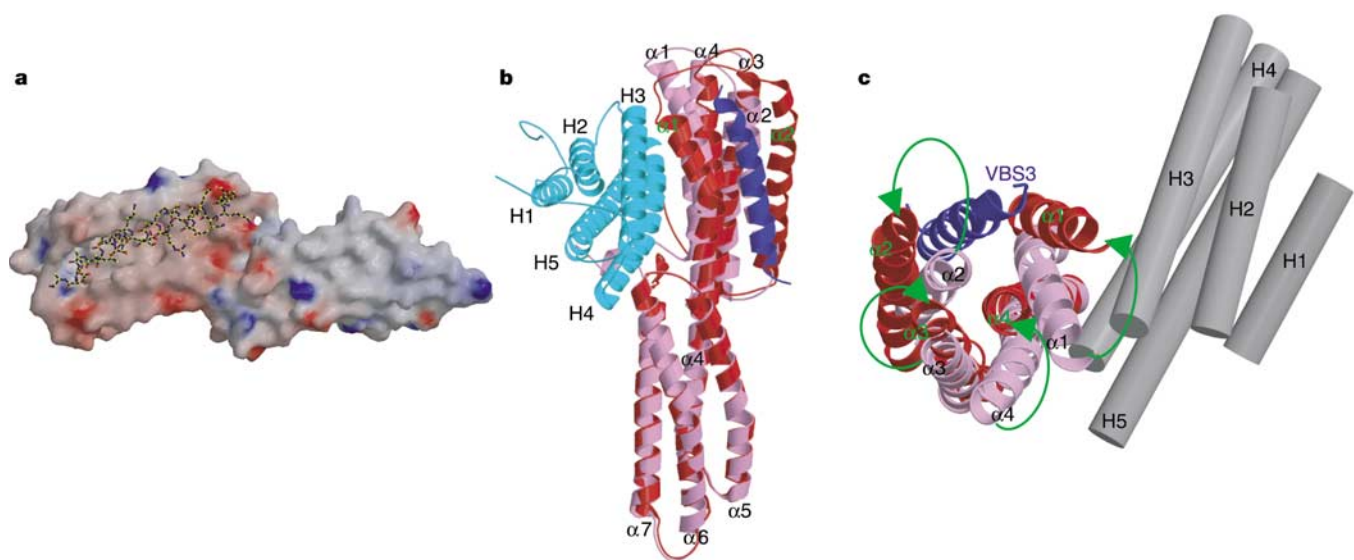


Figure 4 Structure of Vh when activated by talin. **a**, Electrostatic surface potential (red, negative; blue, positive) of Vh when bound to talin. Talin VBS3 is shown in ball-and-stick representation (oxygen atoms, red; carbon, yellow; nitrogen, blue; bonds, black). **b**, The C-terminal bundle (helices $\alpha 4$ –7) of active (red) and inactive (pink) Vh are superimposed

(back view of Fig. 3a). Talin VBS3, dark blue; Vt, light blue. **c**, Movements and helical distortions (green arrows) of the helices ($\alpha 1$ –4) of the N-terminal bundle of inactive Vh (pink) occurring on activation of Vh (red) by talin VBS3 (dark blue). Helices H1–5 of inactive Vt (grey) are shown when bound to Vh (pink).

helical bundle conversion in the N-terminal helical bundle of Vh.

To evaluate whether talin can indeed displace Vt from the Vh–Vt complex, we titrated the VBS3 peptide into pre-existing Vh–Vt complexes. Even at one-third molar ratio, VBS3 displaced Vt to

form a new complex, and higher levels of VBS3 totally disrupted the Vh–Vt complex (Supplementary Fig. S3). The VBS3-binding site in our Vh–VBS3 crystal structure overlaps with the binding site for α -actinin, and both VBS3- and α -actinin-binding sites (residues 713–749) are amphipathic α -helices^{7,11}. Indeed, this α -actinin peptide effectively displaced Vt from Vh to form a new complex, and the binding of α -actinin and VBS3 to Vh was mutually exclusive (Supplementary Fig. S3).

Previous studies have suggested that vinculin activation is largely due to simply severing the Vh–Vt interaction, and that this is triggered by the binding of acidic phospholipids to Vt^{3,10}. In this model Vh was assumed to have a largely passive role, by providing an inflexible scaffold that bound to its partners through helical exchange (Supplementary Fig. S4a), as seen in the binding of β -catenin to α -catenin¹⁴, or by helical addition (Supplementary Fig. S4b), as seen in the binding of paxillin LD motifs to the focal adhesion-targeting domain of FAK^{23,24}. Thus, our finding that half of the α -helices of Vh undergo dramatic movements, rearrangements and distortions on talin binding is completely unexpected. Vinculin can now no longer be viewed as having a rigid head domain with a flexible tail, but is rather revealed as a structurally dynamic protein that moves, bends and flexes to accommodate its diverse binding partners and functions. Importantly, these findings establish helical bundle conversion (Supplementary Fig. S4c) as a new mechanism that alters protein structure.

Talin's ability to provoke marked changes in Vh structure and to displace Vt from pre-existing Vh–Vt complexes supports the concept that talin activates vinculin in focal adhesions. The structures reveal that the talin-binding site does not overlap with that of Vt, but instead that talin distorts the Vh–Vt interface from a distance to displace Vt. Therefore, talin binding may not require unfurling of Vt and might have a direct role in severing the Vh–Vt interaction. In addition, talin binding and activation of PIPKI γ (phosphatidylinositol phosphate kinase γ) leads to increases in PtdInsP₂^{12,13}, underscoring the proximal role for talin as a key trigger for vinculin activation in focal adhesions.

Another insight is that the structural changes in Vh probably serve to direct cytoskeletal assembly in focal adhesions and adherens junctions. Specifically, the alterations of the α -helices within the

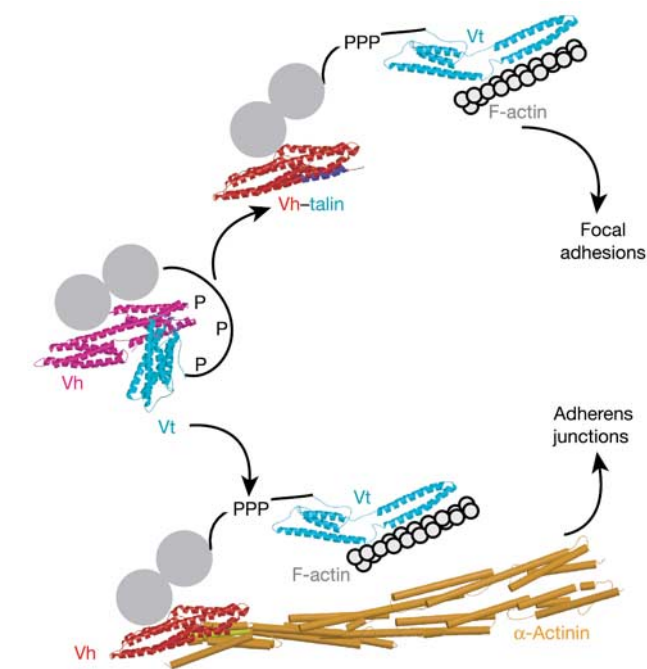


Figure 5 Vinculin activation in focal adhesions and adherens junctions. In its resting state vinculin is kept in a closed conformation through interactions of the N-terminal helical bundle of Vh (pink) with the helical bundle of Vt (light blue). In focal adhesions talin (dark blue) binding provokes helical bundle conversion in Vh, transforming this into an entirely new helical structure (red). Talin binding displaces Vt, allowing it to bind to F-actin and other cytoskeletal proteins. The α -helical-binding domains of talin and α -actinin (yellow) interact with Vh in a mutually exclusive fashion. Thus, in adherens junctions α -actinin binding (orange) triggers different alterations in Vh structure. Grey spheres denote less-conserved globular domains observed by electron microscopy⁴.

N-terminal bundle of Vh establish that the structure of this domain is regulated by contacts with cytoskeletal proteins. For example, the binding of talin between helices $\alpha 1$ and $\alpha 2$ displaces Vt, and α -actinin binding to Vh also displaces Vt, yet talin and α -actinin bind to Vh in a mutually restricted manner. The exclusive interactions of these α -helices with Vh would thus serve to properly direct cytoskeletal and signalling decisions induced by focal adhesions versus adherens junctions (Fig. 5).

The remarkable structural changes in the α -helices of the N-terminal bundle of Vh establish helical bundle conversion as a mechanism of control in signal transduction pathways. Helical bundle conversion is fundamentally different from helical exchange, where the binding of one α -helix simply displaces the binding of another without altering the structure of the displaced helix (Supplementary Fig. S4). Rather, helical bundle conversion provokes dramatic changes in the conformation of α -helices, allowing proteins to expand their repertoire of structures, substrates and functions. □

Methods

Detailed information on methodological aspects of X-ray data collection, structure determination and refinement is provided as Supplementary Information.

Protein preparation

Vinculin complementary DNA (GenBank accession number NM 003373) was generated by polymerase chain reaction with reverse transcription (RT-PCR) of human fibroblast messenger RNA, and was cloned into the pET3 expression vector (Novagen). The final constructs were N-terminal hexahistidine fusion tags proceeded by Met-Glu and included amino acids 1–258 for human Vh and 879–1066 for human Vt. Proteins were expressed in *Escherichia coli* BL21(DE3) or in B834(DE3) for selenomethionine (SeMet) incorporation. Cells were lysed in Tris-HCl (pH 8), 0.5 M NaCl and PMSF, and expressed proteins were purified using a chelating nickel affinity column (Amersham). The protein was eluted over a gradient to 0.5 M imidazole. Vh was purified using an anion-exchange column (Q Sepharose, Amersham), whereas Vt was purified using a cation-exchange column (SP Sepharose, Amersham) and a Superdex 75 gel-filtration column in 200 mM NaCl and 10 mM Tris-acetate (pH 7.6). Proteins were dialysed into Tris-acetate (pH 7.6), 10 mM dithiothreitol (DTT) and 1 mM EDTA. For Vt dialysis buffer also included 150 mM NaCl. Proteins were concentrated to 10 mg ml⁻¹, aliquoted and stored at -20 °C.

Vh-Vt crystallization

Native Vh-Vt crystals and native Vh in complex with SeMet-substituted Vt crystals were obtained by vapour diffusion by equilibrating 1.5 mol Vt per mol of Vh against a reservoir containing 15% polyethylene glycol of molecular mass 3350 Da (PEG-3350), 50 mM NaCl, 100 mM Tris (pH 8), and 10 mM DTT. Crystals of approximate dimensions 0.2 × 0.15 × 0.05 mm³ appeared overnight at room temperature. These crystals belong to space group P2₁2₁2 with one heterodimer in the asymmetric unit, a solvent content of 46%, and a volume to mass ratio of 2.3 Å³ Da⁻¹. Crystals were cryoprotected in 30% PEG-3350.

Vh-VBS3 crystallization

Talin VBS3 (residues 1944–1969) was synthesized and purified by high-performance liquid chromatography in our in-house facility. Initial crystallization conditions were identified at the Hauptman-Woodward Institute. Native and SeMet Vh-VBS3 crystals were obtained by equilibrating 2.4 mol VBS3 per mol Vh against a reservoir of 2% MPD, 100 mM citric acid (pH 4) and 100 mM CdCl₂. Crystals of approximate dimensions of (0.15 mm)³ appeared within one week at room temperature. These crystals belong to space group I4₁32 with one heterodimer in the asymmetric unit, a solvent content of 44% and a volume to mass ratio of 2.2 Å³ Da⁻¹. The Vh-VBS3 crystals were cryoprotected in paratone oil.

Received 28 August; accepted 9 December 2003; doi:10.1038/nature02281.
Published online 31 December 2003.

- Zamir, E. & Geiger, B. Molecular complexity and dynamics of cell-matrix adhesions. *J. Cell Sci.* **114**, 3583–3590 (2001).
- Pokutta, S. & Weiss, W. I. The cytoplasmic face of cell contact sites. *Curr. Opin. Struct. Biol.* **12**, 255–262 (2002).
- Xu, W., Baribault, H. & Adamson, E. D. Vinculin knockout results in heart and brain defects during embryonic development. *Development* **125**, 327–337 (1998).
- Winkler, J., Lunsdorf, H. & Jockusch, B. M. The ultrastructure of chicken gizzard vinculin as visualized by high-resolution electron microscopy. *J. Struct. Biol.* **116**, 270–277 (1996).
- Gilmore, A. P. & Burridge, K. Regulation of vinculin binding to talin and actin by phosphatidylinositol-4-5-bisphosphate. *Nature* **381**, 531–535 (1996).
- Miller, G. J., Dunn, S. D. & Ball, E. H. Interaction of the N- and C-terminal domains of vinculin. Characterization and mapping studies. *J. Biol. Chem.* **276**, 11729–11734 (2001).
- McGregor, A., Blanchard, A. D., Rowe, A. J. & Critchley, D. R. Identification of the vinculin-binding site in the cytoskeletal protein α -actinin. *Biochem. J.* **301**, 225–233 (1994).
- Johnson, R. P. & Craig, S. W. F-actin binding site masked by the intramolecular association of vinculin head and tail domains. *Nature* **373**, 261–264 (1995).

- Wood, C. K., Turner, C. E., Jackson, P. & Critchley, D. R. Characterization of the paxillin-binding site and the C-terminal focal adhesion targeting sequence in vinculin. *J. Cell Sci.* **107**, 709–717 (1994).
- Bakolitsa, C., de Pereda, J. M., Bagshaw, C. R., Critchley, D. R. & Liddington, R. C. Crystal structure of the vinculin tail suggests a pathway for activation. *Cell* **99**, 603–613 (1999).
- Bass, M. D., Smith, B. J., Prigent, S. A. & Critchley, D. R. Talin contains three similar vinculin-binding sites predicted to form an amphipathic helix. *Biochem. J.* **341**, 257–263 (1999).
- Ling, K., Doughman, R. L., Firestone, A. J., Bunce, M. W. & Anderson, R. A. Type I gamma phosphatidylinositol phosphate kinase targets and regulates focal adhesions. *Nature* **420**, 89–93 (2002).
- Di Paolo, G. *et al.* Recruitment and regulation of phosphatidylinositol phosphate kinase type I gamma by the FERM domain of talin. *Nature* **420**, 85–89 (2002).
- Pokutta, S. & Weiss, W. I. Structure of the dimerization and β -catenin-binding region of α -catenin. *Mol. Cell* **5**, 533–543 (2000).
- Yang, J., Dokurno, P., Tonks, N. K. & Barford, D. Crystal structure of the M-fragment of α -catenin: implications for modulation of cell adhesion. *EMBO J.* **20**, 3645–3656 (2001).
- Pokutta, S., Drees, F., Takai, Y., Nelson, W. J. & Weiss, W. I. Biochemical and structural definition of the I-afadin- and actin-binding sites of α -catenin. *J. Biol. Chem.* **277**, 18868–18874 (2002).
- Weiss, E. E., Kroemker, M., Rudiger, A.-H., Jockusch, B. M. & Rudiger, M. Vinculin is part of the cadherin-catenin junctional complex: complex formation between α -catenin and vinculin. *J. Cell Biol.* **131**, 755–764 (1997).
- Watabe-Uchida, M. *et al.* α -catenin-vinculin interaction functions to organize the apical junctional complex in epithelial cells. *J. Cell Biol.* **142**, 847–857 (1998).
- Johnson, R. P. & Craig, S. W. An intramolecular association between the head and tail domains of vinculin modulates talin binding. *J. Biol. Chem.* **269**, 12611–12619 (1994).
- Bass, M. D. *et al.* Further characterization of the interaction between the cytoskeletal proteins talin and vinculin. *Biochem. J.* **362**, 761–768 (2002).
- Lawrence, M. C. & Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950 (1993).
- Johnson, R. P. & Craig, S. W. An intramolecular association between the head and tail domains of vinculin modulates talin binding. *J. Biol. Chem.* **269**, 12611–12619 (1994).
- Hayashi, I., Vuori, K. & Liddington, R. C. The focal adhesion targeting (FAT) region of focal adhesion kinase is a four-helical bundle that binds paxillin. *Nature Struct. Biol.* **9**, 101–106 (2002).
- Hoellerer, M. K. *et al.* Molecular recognition of paxillin LD motifs by focal adhesion targeting domain. *Structure* **11**, 1207–1217 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Cleveland for many helpful discussions. We also thank V. Morris, K. Brown and C. Kirby for technical assistance; C. Vornrhein for expert advice and help with autoSHARP; C. Ross for maintaining the X-ray and computing facilities; L. Messerle for the tantalum compound; and M. Kastan for critical review of the manuscript. We are grateful to the staff at the Advanced Photon Source, COM-CAT, SBC-CAT and SER-CAT, and at the Advanced Light Source, Lawrence Berkeley Laboratory, 5.0.2, for synchrotron support. This work was supported in part by the Cancer Center Support (CORE) Grant and by the American Lebanese Syrian Associated Charities (ALSAC). P.B. is a Van Vleet Fellow.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.I. (tina.izard@stjude.org). The atomic coordinates and structure factors have been deposited in the Protein Data Bank under accession codes 1RKE and 1RKC for the Vh-Vt and Vh-VBS3 structures, respectively.

erratum

Hydrocarbons and the evolution of human culture

Charles Hall, Pradeep Tharakan, John Hallock, Cutler Cleveland & Michael Jefferson

Nature **426**, 318–322 (2003).

In Table 1 of this Insight, the USGS estimate of world oil ultimate recovery should have cited ref. 10, not ref. 11. References 10 and 11 should read as below^{10,11}. In addition, in ref. 17, P. Tharakan's surname was misspelt as Tharkan. □

- United States Geological Survey (USGS) *U.S. Geological Survey World Petroleum Assessment 2000 - Description and Results*. Version 1.1 (USGS Digital Data Series DDS-60, United States Geological Survey, Washington DC, 2000).
- Energy Information Administration, United States Department of Energy (EIA/DOE) *Long-term World Oil Supply: A Resource Base/Production Path Analysis* (US DOE, Washington DC, 2000); at http://tonto.eia.doe.gov/FTPROOT/presentations/long_term_supply/index.htm.

Contacts

Publisher: Ben Crowe
Editor: Paul Smaglik
Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs Sales Director:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)
Andy Douglas (4975)
Frank Phelan (4944)
**Netherlands/ Italy/ Spain/
Portugal/ Belgium:**
Evelina Rubio Hakansson (4973)
Scandinavia: Silje Opstrup (4994)
France/ Switzerland:
Amelie Pequignot (4974)
Natureevents:
Paul Constant (4954)

Production Manager:

Billie Franklin
To send materials use London
address above.
Tel +44 (0) 20 7843 4814
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen
Tel + 49 89 54 90 57 11/-2
Fax + 49 89 54 90 57 20
e-mail: p.phelan@nature.com
o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718
Fax +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager:

Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841
Tel +81 3 3267 8751
Fax +81 3 3267 8746
Asia-Pacific Sales Director:
Rinoko Asami
e-mail: rasami@naturejpn.com

naturejobs

Tuning in to yourself

When I've talked to scientists about how they got to where they are, their journeys often turn out to have been surprisingly undirected. When Ravinder Maini (see page 180) was earning his MD, for example, I doubt that he set himself the goal of discovering an antibody that would help to treat autoimmune diseases and win him a Lasker award. But he followed his interests, intuition and the advice of his mentor, and went into clinical research rather than clinical practice.

This year *Naturejobs* is asking four graduates to keep a formal career journal that notes what happens to them day by day — not to ensure that they win a Lasker award, but rather to tune into how their likes and dislikes could shape the kind of career they want. The idea, according to Deb Koen, who writes about how to keep a career journal in this week's Nuts & Bolts (see page 180), is to become self-aware enough to find the best career path for you.

At three different career-panel discussions recently — in London, Boston and Connecticut — at least one panellist has encouraged students and postdocs to “do what you like”. But that isn't always easy, says Koen. Scientists tend to think of workplace preferences in terms of research — a penchant for signal transduction over genetics, or developmental biology over oncology. Analysing your own likes and dislikes throughout the day may be a better guide, because although scientific problems change, personalities are relatively fixed.

Koen asked the four diarists to start with basic questions such as: what do you spend your time on each day, and what do you like and dislike about each aspect of these tasks? Such questions are not limited to grad-school diarists — keeping them in mind can help even established scientists make a change for the better.

Paul Smaglik

Naturejobs editor



Contents

POSTDOCS AND STUDENTS

The wages game p178

CAREER VIEW

Nuts & Bolts

Keeping a career diary

Graduate Journal

A change of direction

Movers

Ravinder Maini p180

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Healthy limits

The US National Institutes of Health stipends are being used as de facto guidelines for postdoc salaries. But do institutions try to meet them? Karen Kreeger investigates.

Is there a minimum wage for US postdocs? If so, who sets it and who enforces it? The answer to the first question is: “sort of” — which is somewhat explained in the answer to the second. The salary levels of the National Research Service Awards (NRSA) given by the US National Institutes of Health (NIH) are being used as de facto guidelines by postdocs and their supporters in university administration, in seeking pay rises. This practice has become more widespread since 2001, when the NIH bumped up the minimum NRSA salary levels by several thousand dollars for all levels of postdoctoral experience. In March 2003, NRSA stipend levels for entry-level postdocs rose to \$34,200, from \$31,092 the previous year.

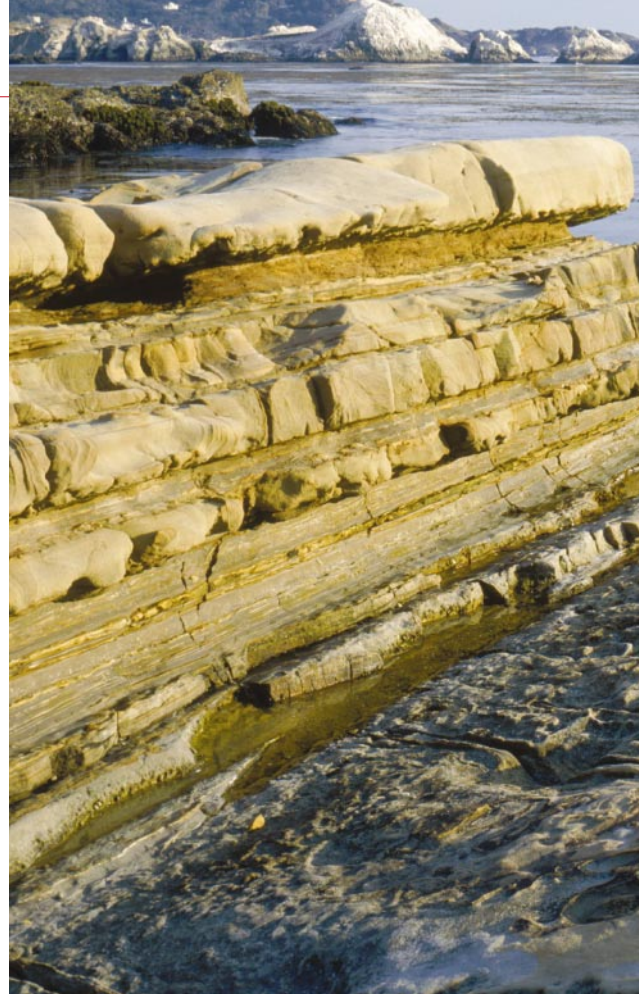
Administrators both inside and outside the United States take note of the NRSA scales, but these are not official guidelines and have no teeth. Stipends still vary from institution to institution, even if their recipients have similar skills and experience. Some universities have started comparing their postdoc salaries with those offered by other institutions, and some have established guidelines of their own. Still, the NRSA scale remains the most widely used criterion.

The NIH receives many queries from researchers and administrators asking about salary standards. These are fielded by Walter Schaffer, acting director of the NIH office of extramural programmes and a research training officer, who states unequivocally that the NIH simply sets stipend levels for the NRSA programme alone — it doesn’t publish standards for salaries. Although there are no exact statistics of how many institutions use the NRSA levels as a guideline, Schaffer feels that they are pervasively used in that way. “There’s really nothing we can do. We publish the stipends and it clearly applies to the NRSA programme,” he says.

Low wages are “a national issue that is being faced by virtually every university, medical school and research institution that ‘employs’ postdoctoral fellows,” says Roslyn Orkin, assistant dean for faculty affairs at Harvard Medical School.

MEETING AND EXCEEDING

Many institutions are trying to meet or exceed the NRSA levels. According to Christine Hickey, spokeswoman for the Memorial Sloan-Kettering Cancer Center in New York, postdoc salaries there have always been higher than the NIH minimum because of the high cost of living in Manhattan. “In addition, we provide some funding for each lab budget so that the burden does not fall squarely on the shoulders of the principal investigator alone,” Hickey says. At Harvard, the preclinical departments have strongly suggested



The National Institutes of Health NRSA stipend is used by many as a benchmark. Although the agency aims to raise the entry-level figure to \$45,000, there is no guarantee that the rising tide will lift everybody's boat.

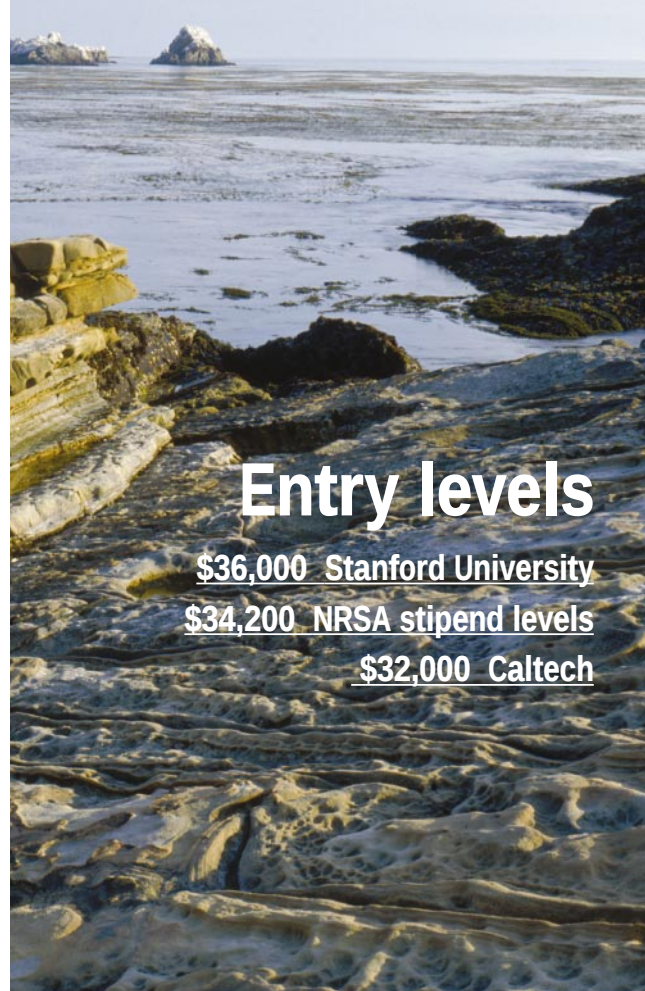
J. FOOT/NATUREPL

that, at least for the first two years of a postdoc, faculty members should better the NRSA minimum rates by \$2,000. “To my knowledge, individual principal investigators and departments have been able to meet this financial challenge,” says Orkin.

A few universities publish salary guidelines on the web. For example, Rockefeller University publishes a range, starting from \$34,000–\$39,000 for less than one year of experience. At Stanford University, says Karen Christopherson, a postdoc in neurobiology there, the first years of a postdoctoral fellowship used to be below the NIH numbers, but now entry-level postdoc salaries exceed the NRSA minimums. For autumn 2003, Stanford’s minimum was \$36,000.

“We have a postdoc office with a dean who takes this very seriously and is making sure that everyone is making the minimum,” Christopherson says. “I think we were successful because we had supportive faculty and administration in place that wanted to push this through and we have a provost who stood very strong on this issue.” John Boothroyd, professor of microbiology and immunology, and senior associate dean for research at Stanford University School of Medicine, says that the vast majority of faculty members recognize that raising salaries is the right thing to do. “The rate before for postdocs has been embarrassingly low,” he says.

At Rockefeller most of the initiative has come from the administration, notes Kevin O’Donovan, a postdoc in the laboratory of molecular neuro-oncology, who is



Entry levels

\$36,000 Stanford University
\$34,200 NRSA stipend levels
\$32,000 Caltech

active in its postdoctoral association. Following a decree by the university's former president Arnold Levine, salaries for Rockefeller postdocs have increased twice in the past two years to stay ahead of the NRSA minimum rate. Maria Lazzaro, director of immigration and academic appointments, sent e-mails and letters to laboratory heads to ensure that they complied.

"We met with the postdoc association informing them of the new guidelines," says Lazzaro. Then the department followed up by going through each person's file to make sure they're being paid the correct amount, Lazzaro says. Rockefeller's postdoc association also e-mailed all the postdocs and lab heads about the salary increase.

NOT RETROACTIVE

While postdocs rejoiced, the pay rises have caused difficulties for some researchers who have to meet them out of shrinking grants. The effect of increases on lab budgets is not to be taken lightly, say administrators. "In every single area we are at or well above the minimum," says Amy Gutmann, provost of Princeton University. "The faculty members have been very supportive in the cases where the minimum has increased the floor."

Administrators agree that it is especially difficult if researchers have existing NIH grants. "The reaction from faculty members has been in favour, but some have budgetary constraints," says Rockefeller's Lazzaro. "One thing that has come to my attention is that faculty NIH grants are for five years. They get the budget they asked for years ago and it doesn't have any built-in cost-of-living increase. So the NIH on one hand is encouraging certain salary levels for postdocs,

POSTDOCS & STUDENTS

and so are we, but labs that rely on those grants — particularly those headed by younger principal investigators — don't have the NIH funding to cover postdoc salary increases over time."

Helen McBride, a postdoctoral scholar in developmental biology at the Beckman Institute in Pasadena, part of the California Institute of Technology (Caltech), concurs. She thinks these salary increases could affect junior faculty members more because they generally don't yet have an array of grants to supplement salaries. But Boothroyd counters that new faculty members get start-up funds and are less reliant on NIH money than more senior researchers.

Despite most institutions' efforts to abide by the suggested minima, "no guidelines for implementation have been established and NIH increases are not retroactive", says Orkin. "Thus, although the NIH stipend scale is marching upwards, there is no funding mechanism to provide similar percentage increases to those whose funding source began *x* years ago, when the starting stipends were lower."

PLAYING CATCH-UP

Other universities are still trying to gather information on the range of postdoc salaries within their own institutions. Jim Randolph, senior associate director of research administration at the University of Michigan, Ann Arbor, says that there has "most decidedly" been an increase in Michigan postdocs' pay over the past few years, in line with increases by the NIH, whose goal is to raise stipends by 10–12% per year until entry-level postdocs start at \$45,000.

Christopher Hall, a research fellow in the department of pathology at the University of Michigan, says that the Michigan postdoc association has yet to take on salaries as an issue. The postdocs hope to address the issue after a survey by the scientific research society Sigma Xi has been carried out next spring, when they will have a better measure of how their salary levels stack up nationally.

Randolph says that most postdocs he knows are not earning below NRSA levels. "I have to think that we do pretty well here, but it's hard for me to generalize. We really don't talk salaries," he says. He recently started a second postdoc and used the NRSA criteria to gauge what salary level to ask for in interviews.

Despite good intentions, not all universities have been able to raise the bar. "If you want to change things and bring the whole scale up to the NRSA minimum it has to be a grassroots faculty initiative," says Caltech's McBride. "We're trying to build that support now." Caltech has instituted a pay freeze for faculty and staff who make over \$50,000, but staff under that level, including postdocs, will see a 3% rise.

The Caltech postdoc association has written to faculty members seeking their concerns about raising postdoc salaries and asking how this would affect each researcher. "We got the response we pretty much expected, which was: 'Hey, it's expensive to run a lab,'" says McBride. "As someone going out to start a lab and fully aware of how much it costs to pay people a decent wage, this means my lab will stay small and I don't see anything wrong with that." Many institutions are well on their way to paying postdocs a decent wage, but the road to equity is still a long one. ■

Karen Kreeger is a freelance science writer in Media, Pennsylvania.

Web links

NRSA stipend levels

► grants.nih.gov/training/nrsa.htm

Rockefeller University salary guidelines

► www.rockefeller.edu/pda/PDAnews.html

Stanford University salary guidelines

► postdocs.stanford.edu/handbook/salary.html

University of Michigan

Postdoctoral Association

► www.med.umich.edu/pibs/postdoc/aboutus.html

Caltech Postdoctoral Association

► cpa.caltech.edu

Johns Hopkins School of Medicine Postdoctoral Association

► www.hopkinsmedicine.org/jhpda/index.html

GRADUATE JOURNAL

From biochemist to engineer

When I was an undergraduate, I sculled my way into a new major — and onto a different career path. When I matriculated into McGill University, I started off in biochemistry and joined the rowing team. Two chemical-engineering student rowers introduced me to their programme during trips to competitions, and I decided to transfer to engineering.

After graduating, I worked as a junior process engineer for an engineering consultancy. A year later I returned to McGill for a master's degree, aiming to specialize in crystallization and learn project management. I was enthralled by the experience, but returned to industry.

A year later, I began to miss figuring out new phenomena, testing ideas in the lab and tutoring undergraduate students. I reflected on a paper I stumbled across during my master's that discussed the fabrication of ceramic bone biomaterials. I wondered if I could apply my knowledge to designing biomaterials for bone. With the reported mobility issues of the ageing population, I thought it would be motivating and inspiring to join a research team in this field. I was not sure if I could handle a second, longer, pay cut. I was more sure that I would regret it if I did not give it a try.

Sidney Omelon is a PhD student in bone biomaterials at the Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Canada.

NUTS & BOLTS

Career journals

Reflecting on your career tends to take a back seat to demanding schedules and work priorities. In the New Year there's no better way to remedy this than by keeping a career journal. This takes only a little time each month, yet the rewards are tenfold in career focus and effectiveness.

Whether you're a student, postdoc or mid-career scientist, making regular diary entries will give you an opportunity to reflect on your current state and future direction. And because your journal is intended for your eyes only, you don't have to worry about being judged on your career goals or your writing skills. If you've made your resolutions for the New Year, a journal will keep you on track. If you're lacking career direction, it will help you focus.

The discipline of regular writing will prompt you to



With Deb Koen
Careers consultant

clarify your strengths and passions, identify gaps in expertise and tune in to personal development needs. By tracking progress over time, the journal will help you learn from mistakes and build on your strengths. The act of writing will solidify your thinking and serve as a catalyst for action.

Choose a notebook or create a PC file that will be immediately accessible when you want to capture a fleeting idea, need an outlet for your frustrations or have a reaction to a recent event. In addition to notes that you jot down

spontaneously, schedule a 'journal hour' to recap each month.

Take a varied approach to writing. Start the hour with 15 minutes of uninterrupted and uncensored freestyle writing. Record whatever enters your mind about your career. Don't stop.

For a more structured recording, select a career-related theme and write away for 15–45 minutes. Topics might include peak moments at work, family influences on your career, work-life balance or lack of it, your sense of 'fit' in the organizational culture, potential career derailers, priority development needs, or trends most likely to have an impact on your career.

Now pull out a pen and notebook, or pull up a chair to your computer and start documenting the unfolding drama of your career.

Deb Koen is vice-president of Career Development Services and a columnist for *The Wall Street Journal's CareerJournal.com*.

MOVERS

Ravinder Maini, board member, Domantis, Cambridge, UK



For Ravinder Maini, a mentor set him on the path that would eventually lead to the research that won him and a colleague a Lasker award last year. The molecular-biology revolution helped him complete it.

John Butterfield at Guy's Hospital in London first taught Maini not to accept any assumptions about disease. He remembers Butterfield saying

sardonically that insulin therapy "had put back research on diabetes" because it treated the symptoms — and thus had stifled research into the underlying mechanisms. That kind of critical thinking steered Maini towards research, rather than clinical practice. "I was inquisitive about biological mechanisms of disease," he says.

In 1968 he started work under Dudley Dumonde, one of the first in the United Kingdom to do cytokine research. His group eventually landed a grant to study cytokine expression in rheumatoid arthritis. "That's really where the course of my career changed dramatically," Maini says. But to develop his ideas he had to wait for technology to catch up.

It did when Greg Winter, co-founder of Domantis, humanized monoclonal antibodies and started using them to treat disease. And once the cytokine genes that interested Maini were cloned in the 1980s, he and his collaborator

Marc Feldmann realized that they could perhaps use monoclonal antibodies to treat rheumatoid arthritis and autoimmune diseases. These developments "opened the door", says Maini.

Now that the drugs he helped to develop have proved successful, Maini wants to make them more widely available. The cost of manufacturing therapeutic proteins means that a year's treatment can cost about \$15,000. So when Winter approached Maini to become involved with Domantis, he was eager, as the company's technology promises a cheaper and more specific way to make the proteins for treating the same autoimmune diseases.

Maini has never regretted pursuing clinical research — despite the 30-year-slog from concept to market. "Positive feedback keeps you going personally," Maini says — despite "considerable scepticism" at some points.

CV 2002: Emeritus professor of rheumatology, Imperial College, London; trustee, Kennedy Institute of

Rheumatology, London; and emeritus honorary consultant physician at Charing Cross Hospital, London

2000–2002: Head, Kennedy Institute of Rheumatology; Faculty of Medicine, Imperial College, London

1989–2002: Professor of rheumatology, Imperial College; head, department of rheumatology at Charing Cross Hospital, London

1979–2002: Head of the division of clinical immunology, Kennedy Institute of Rheumatology